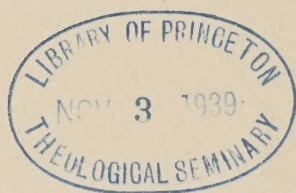


Motivation
of
Young Children

Lucile Chase

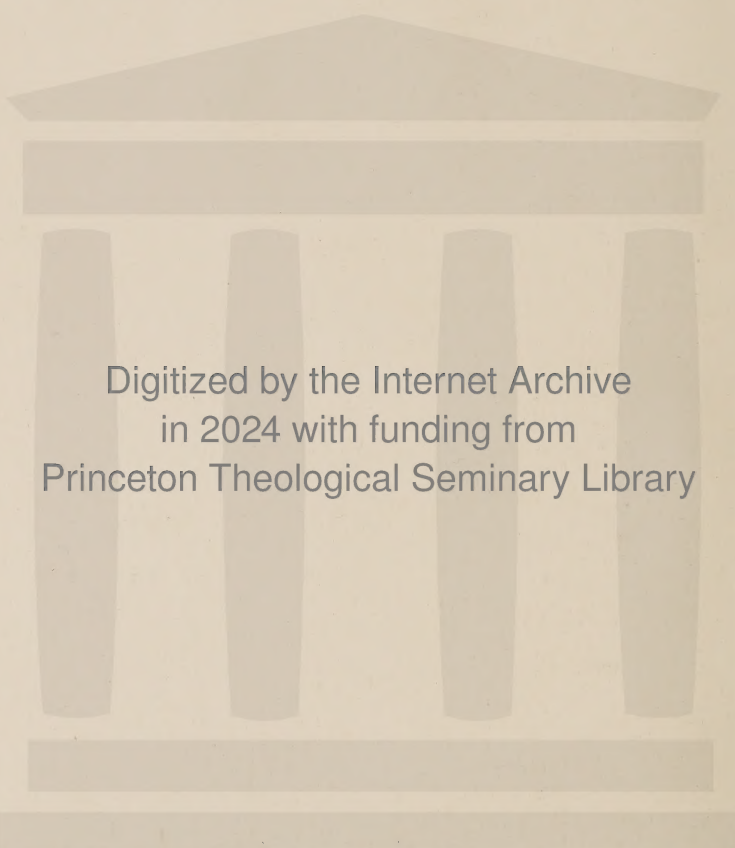
LB1105
.I64
v. 5:3



LB1105

. I64 .

v. 5:3

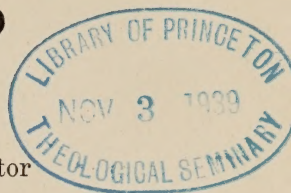


Digitized by the Internet Archive
in 2024 with funding from
Princeton Theological Seminary Library

W
Iowa. University. Studies in child welfare.

UNIVERSITY OF IOWA STUDIES IN CHILD WELFARE

Professor GEORGE D. STODDARD, Ph.D., Editor



FROM THE IOWA CHILD WELFARE RESEARCH STATION

VOLUME V

NUMBER 3

MOTIVATION OF YOUNG CHILDREN

An Experimental Study of the Influence of Certain
Types of External Incentives Upon the
Performance of a Task

by

✓
LUCILE CHASE, Ph.D.

Published by the University, Iowa City, Iowa

1932

ACKNOWLEDGMENT

This study was begun when the writer was a National Research Council Scholar during 1927-1928. She wishes to express her appreciation to Dr. George D. Stoddard for the helpful criticism and assistance which made this investigation possible. She also wishes to thank Mr. M. G. Davis, Superintendent of the Ames Public Schools, for permission to use as subjects children in four kindergartens of that city. To Mrs. Lulu R. Lancaster acknowledgment is due for making available the preschool children in the nursery school at Iowa State College. The writer also wishes to acknowledge the many helpful suggestions which have been given by the members of the staff of the Iowa Child Welfare Research Station.

FOREWORD

At a time when the doctrine of punishment is being seriously questioned, a study of the effectiveness of various types of incentives is especially appropriate.

Experimental work in this problem has long been handicapped by the number of variables to be brought under control and by a lack of suitable apparatus for securing quantitative data. Both these situations have been met successfully by Dr. Chase. She has carefully defined and administered the incentives selected and has devised apparatus well adapted to the measurement of their effects on young children.

For a sampling of 259 children in the age range of two to eight years it appears that the absence of external incentive is consistently a deterrent to improvement in performance. Even with knowledge of success on the part of the child it is likely that praise or reward in addition will be helpful. Similarly, knowledge of failure is better than ignorance of results, but the addition of reproof or punishment is stimulating, with a slight margin in favor of reproof.

In short, the evidence points to a child need not so much for a specific incentive as for *some* incentive above that of knowledge of results.

GEORGE D. STODDARD

Office of the Director
Iowa Child Welfare Research Station
University of Iowa
September 8, 1931

CONTENTS

Chapter	Page
I The Importance and Significance of Motivation	9
II Previous Studies on the Influence of Incentives	15
Effectiveness of Reward and Punishment	15
Relative Effectiveness of Knowledge of Versus No Knowledge of Progress and Results of Performance	18
Influence of Praise and Reproof Upon Performance	21
Effect of Purpose and Aim	23
Influence of Attitude Upon Test Performance	25
Influence of Success and Failure Upon Performance	27
Summary of the Characteristic Findings in the Field	27
Problems in Motivation of Young Children	29
III Statement of the Problem, Method and	
Subjects of the Investigation	31
Statement of the Problem	31
Method	32
Forming Comparable Groups	33
Plan of Testing	33
First Week	34
Second Week	35
Third week	35
Subjects	35
IV Apparatus, Materials, and Test Procedure Used in the Study	37
Motivation-Dynamometer Test	37
Main System	38
Basic System	39
System Maintaining Liquid at Highest Point	43
System Lowering Liquid	44
Lever System	44
System for Insuring Success and Failure	45
Adjustable Chair and Arm Rest	47
Chair	47
Arm Rest	47
Train System	49
Disadvantages	52
Perforation Test	52
Materials of the Reward and Punishment Series	53
Disadvantages	54
Records	55
Instructions to Recorder	55
Behavior Responses to the Experimental Situation	58
Behavior Responses to Different Motivating Conditions	59
Test Procedure	60

Chapter	Page
Method of Obtaining Subject from His Group	60
Adjustment of Subject to Apparatus	
During the Pre-Experimental Period	61
Method of Obtaining Cooperation	61
Specific Technique for Adjusting Subject to Apparatus	62
Directions and Procedure in Giving the	
Three Experimental Test Presentations	63
Presentation 1	63
Presentation 2	65
Presentation 3	67
Procedure During Perforation Test	68
V Analysis of Experimental Results	71
Comparability of Groups	71
Effectiveness of Incentives on Motivation-	
Dynamometer Test	75
Series II	75
Series III	76
Series I and II	77
Series I and III	79
Series II and III	79
Interpretation of Results	79
Series II	79
Series III	82
Series I and II	83
Series I and III	83
Series II and III	83
Variability	84
Influence of Incentives in Maintaining Efficiency	85
Interpretation of Results	92
Qualitative Aspects of the Influence of Motivation	93
Reliability of Recorders	93
Expressive Behavior	93
Subjects Liking Game I or II	95
Attention	96
Mood	98
Age and Sex Differences	99
Reliability of the Motivation-Dynamometer Test	102
Interpretation of Results	103
Influence of Incentives Upon Performance	
in Perforation Test	103
Interpretation of Results	107
Reliability of the Perforation Test	107
Interpretation of Results	108
VI Summary, Results, and Conclusions	109
References	116

CHAPTER I

THE IMPORTANCE AND SIGNIFICANCE OF MOTIVATION

Child psychology is interested not only in describing what a child does and how he does it, but also in explaining and in ultimately predicting why he takes or will tend to take a particular course of action rather than some other. The underlying causes of conduct are of importance in the explanation of behavior problems and are the bases upon which effective remedial treatment is founded. It may be pointed out that behavior problems have assumed great importance in the early years of life because there is some indication that the reconditioning process is accomplished more rapidly and easily at this time and with better effects for the individual himself as well as for society. The term motivation is the "action of the verb motive" which comes from the Latin word *movere*, meaning "to move." A motive, according to the Oxford Dictionary, is that which moves or induces a person to act in a certain way or tends to move a person to a particular course of action. Thus, if we are to be at all successful in the prediction of behavior or in the determination of the underlying causes of the "whys or movers" of conduct, we must be ultimately concerned with the problem of motivation.

The present study is an inquiry into the nature of certain of these "movers of conduct," with particular reference to young children.

In reading the psychology of motivation, one finds a number of terms used more or less interchangeably, as for example: incentive, motive, drive, urge, desire, etc. To avoid confusion it would be well at this point to attempt a clarification of the terms as they will be used in this study.

English (15) in his *Dictionary of Psychological Terms* gives the following definition: "Motivation is that which, from the standpoint of the *reacting organism*, accounts for what one does. No single term is in use for these 'springs of action.' . . . Motivation seems the most general term, not only for the topic but for that which, in conjunction with the *stimulus*, brings about a particular act . . . Motive strongly suggests *conscious purpose* and will." In

differentiating motive from incentive English points out that "incentive is a less general term than motive which it closely resembles. Motives seem to spring from within, incentives from without, though they work from within. An incentive is a sort of indirect or accessory external stimulus."¹

Woodworth (54) states that "in general, a motive is a tendency towards a certain end-result or end-reaction, a tendency which is itself aroused by some stimulus, and which persists for a time because its end-reaction is not at once made. The end-reaction is not made at once because it can only be aroused by an appropriate stimulus, acting in conjunction with the motive. But the motive, persisting in its inner activity, facilitates reactions to certain stimuli and inhibits others."²

Although, from this discussion it is evident that motivation may be concerned with all the "movers" of conduct, the types used in this study should be classed as incentives in that they are external, indirect or accessory stimuli and "spring from without, though they work from within." Certain difficulties would rapidly become apparent if one were to use the term "incentive" consistently, such as, for example, using it as a transitive verb or as a modifying adjective. We do not say "he was incentivated; the incentivation was good; or the incentivated child is the happy child."

This study will not be concerned with those types of motivation designated as drives, urges, desires, needs, but with the more or less external types such as rewards and punishments, praise and reproof, knowledge of progress and results. Motivation and motive will also be used interchangeably with incentive. It seems justifiable to do this since, (1) there is no single term for the "main springs of action"; (2) motivation is the most general term; (3) in addition motive closely resembles "incentive"; (4) unfortunately, there is an idiomatic difficulty in using the term "incentive" consistently.

Motivation has significance in at least three phases of the child's development: (1) as a means of social control; (2) in the formation of habits and in the acquisition of skills and knowledge; and (3) in building up a satisfactory attitude toward the facing of reality.

Motivation as a means of social control is of importance in helping the child to learn to adjust to the group and to authority. Wick-

¹ English, Horace B.: *The Student's Dictionary of Psychological Terms*. Yellow Springs, Ohio: Antioch Press, 1928. Pp. [79]. [p. 40]

² Woodworth, Robert S.: *Psychology: A Study of Mental Life*. New York: Henry Holt, 1921. Pp. x, 580. (p. 84-85)

man (53) states that "the necessity for modifying personal desires according to imposed requirements begins at birth (possibly even before) and becomes increasingly exacting as the individual develops from infancy to adulthood."³ However, Blanton (6) points out that "no king of the ancient world, not even one who considered himself to be a god, was ever cared for as is the infant. For the first fifteen or eighteen months of life, his every want is satisfied . . . As the child emerges from the helplessness of infancy, he comes in conflict with the desires and needs of his brothers, sisters, parents, and other adults who may surround him. His individual desires and wants which conflict with the demands of the group must be adjusted."⁴

If the question is approached from the child's point of view it is found that he may not be willing to relinquish the center of attention without more or less of a struggle. The reluctance to leave the habits of infancy behind may be seen in his attempts to prolong this period in which he is able to exact a large amount of service and attention, by indulging in temper tantrums, in refusal of food, in negativism, in order to preserve in some measure his infantile status of domination over his parents.

Since, as has been suggested, one of the aims of the socializing process is to shape the child's behavior in harmony with the demands of the group and since the child may not submit to these demands willingly, the problem of parent, educator, or whoever is responsible for the child's welfare may be stated in the form of a question: "How can I best influence this child's behavior so that it will be in harmony with the demands of the group?" Or "How can I influence the child to want to do what is necessary for him to do?" According to Thomson (47) "whenever a person aims to influence the conduct of another he must, if he is to succeed, arouse some one or a group of impulses, incentives, urges, drives, motives."⁵ That a large proportion of those people responsible for the child's welfare have not been successful in appealing to the right incentives or

³ Wickman, E. K.: *Children's Behavior and Teachers' Attitudes*. New York: Commonwealth Fund, Division of Publications, 1928. Pp. 247. (p. 133)

⁴ Blanton, Smiley, and Blanton, Margaret Gray: *Child Guidance*. New York: Century, 1927. Pp. xviii, 301. (p. 164)

⁵ Scott, Walter Dill: *Influencing Men in Business: The Psychology of Argument and Suggestion*. 2nd ed. New York: Ronald Press, Pp. 186. (p. 25). Cited by Thomson, Mehran K.: *The Springs of Human Actions: A Psychological Study of the Sources, Mechanism, and Principles of Motivation in Human Behavior*. New York: Appleton, 1927. Pp. xvi, 501. (p. 1)

motives in their attempt to influence conduct is evidenced by the high incidence of undesirable behavior in public school children and in the number of more advanced cases which are referred to the clinics. Blanchard (5) states that "More than half the children now growing up show behavior patterns which indicate that socialization has failed at certain stages of their development."⁶ This statement has been verified by the statistical studies of Haggerty (21) and Wickman (53). Blanchard goes on to say that the outstanding conditions in the home causing these problems are poor training and discipline, which, it may be added, are due in the final analysis to inadequate or unsatisfactory motivation on the part of those who are responsible for influencing the conduct of the children concerned.

The significance of motivation in the acquisition of skills and in habit formation in young children is assumed by those who contend that the fixation and elimination of acts must be explained, in part, in terms of their affective consequences. Thorndike's (49) law of effect is stated: "What comes after a connection acts upon it to alter its strength."⁷ Thorndike has reported, in the *Proceedings and Papers of the Ninth International Congress of Psychology*, two articles bearing on this problem. If by his statement (50) "annoying or painful consequences do not weaken the connection to which they are attached but only strengthen some other connection" he means that a connection is indirectly weakened by punishment, it would seem that he had not fundamentally revised his original position.

If by the statement (51) "a wrong or annoying response facilitates consummation or freedom for further activity or completeness as much as a right response in every sense except that it is not as satisfying" the author means that annoyingness stamps in a connection, then it would seem that he has contradicted his original position. These two inferences drawn by this writer may not be valid because of the condensed and obscure nature of the two reports. The controversy as to whether the accruing emotional satisfaction or dissatisfaction actually is or is not a factor in fixating and eliminating responses or acts will not be reviewed here. It is of interest, however, that so much importance is attached to this

⁶ Blanchard, Phyllis M.: *The Child and Society: An Introduction to the Social Psychology of the Child*. New York: Longmans, Green, 1928. Pp. xi, 569. (p. 268)

⁷ Thorndike, Edward L.: *The Law of Effect*. *Amer. J. Psychol.*, 1927, 39, 212-222. (p. 212)

concept of emotional satisfaction. Educators and mental hygienists use it so prevalently in the explanation of habit formation and as a basis for the remedial treatment of the maladjusted child. The tendency to emphasize this approach is shown in the writings of Blatz (7), Freeman (16), Kilpatrick (31), Lowrey (34), Thom (46), and others.

In addition to this more intrinsic relationship between motivation and habit formation, there are other more extrinsic types used in helping the child to attain some new habit or skill; for instance, it is often necessary to use some incentive in helping a child to initiate a new course of action, in enabling him to overcome fear of a strange situation, and in giving him a feeling of confidence and security. To emphasize the desired or correct response or behavior, and to help him increase or maintain his efficiency are other uses of motivation.

The problem of facing reality frankly, of being able to accept defeat as well as success with a sane and well-balanced emotional attitude becomes of paramount importance in the child's total personality development. It is evident that the child cannot always be successful in what he sets out to do. There may and perhaps often will be success, but most often there will be only partial success and many times failure. A child who can meet the situations of life squarely, who can take defeat and success as "part of the game" has little need to habitually rationalize, lie, steal, run away, resort to temper tantrums, or do a hundred and one other things that are classed as "undesirable behavior." The need for behavior and guidance clinics might not be so pressing if all, rather than only half of the children now growing up, could face reality frankly.

One factor in the development of a sane and well-balanced attitude toward reality is the type of experience which a child has had with success and failure, praise and reproof, reward and punishment. Maladjustment may follow either the type of motivational experience emphasizing success and accomplishment or that which constantly stresses failure and shortcomings. It is probable that a child who has had practice in meeting a more nearly equal amount of both types of motivation will learn to accept defeat or success with greater frankness.

If motivation is of such importance and significance in the life of the child our next question is what types or kinds of motivation will best bring about social control, better performance of a task, or an

adequate adjustment toward reality. As our next chapter will point out, we have, at present, very little experimental evidence with young children indicating the effectiveness of different incentives, even of those in more common use such as praise, reproof, punishment, or reward. It is the purpose of this investigation to obtain data on the relative effectiveness of a number of different types of incentives which have been used under controlled and experimental conditions with four comparable groups of children ranging in age from two to eight years.

CHAPTER II

PREVIOUS STUDIES ON THE INFLUENCE OF INCENTIVES

The previous chapter has pointed out the importance and significance of motivation, more generally as a basis of explanation in child psychology and more specifically as it concerns at least three phases of development in the life of the young child. It was suggested that we should find very few experimental investigations on the influence of motivation using young children as subjects. While this is true, there have been several important studies on the effectiveness of incentives using as subjects adults and older children as well as animals. These investigations will be reviewed under the following heads and it is hoped will serve as a means of orienting the results of the present study: (1) the relative effectiveness of reward and punishment; (2) the relative effectiveness of knowledge of progress and results of performance; (3) the influence of praise and reproof; (4) the influence of attitude upon performance; (5) the influence of success and failure; (6) the effectiveness of purpose and aim. In addition to this review, a summary of the characteristic findings in the field will be indicated, together with a statement of some problems which arise for investigation with young children.

EFFECTIVENESS OF REWARD AND PUNISHMENT

In 1907 Yerkes (56) put forth the theory that the "method of punishment is more satisfactory than that of reward as a method of motivation." Accordingly in 1908 Yerkes, working with Dodson (13), investigated the effect of different intensities of electric shock upon the formation of the white-black discrimination in the dancing mouse. Different degrees of difficulty (easy, medium, and difficult) and three strengths of stimuli (weak, medium, and strong) were used. It was found that the relation of the strength of the electrical stimulus to the rapidity of the learning depended upon the difficulty of the acquisition of the habit. The greater the differences between the discrimination, the greater the beneficial effect on learning, as the strength of the shock was increased and vice versa. Cole (12) tested chicks for the Yerkes and Dodson law and found that it applied in the essentials though not in all details.

The effectiveness of punishment (electric shock) and reward (food) were studied by Hoge and Stocking (24) in 1912. They used as a measure of effectiveness the ability of six female rats to discriminate two different intensities of light as given under the following conditions: The rats were divided into three groups, two to a group, the first group being given punishment (electric shock) for incorrect responses and reward (food) for the correct; the second group being punished for incorrect responses alone; and the third group being rewarded for correct responses alone. The authors concluded as a result of their investigation that it seemed evident that a combination of punishment and reward motives was more effective in bringing about visual discrimination in the rat than was either punishment or reward used alone. It seemed evident also, that punishment is more effective than reward at least in so far as the rate of learning was concerned.

In 1917 Dodson (13) used eighty rats in his study on the relative values of reward and punishment. The rats were required to discriminate between a dark and a light box in order to reach their nest. If the wrong box were chosen an electric shock was given; if the right box were entered the rat was allowed to continue to his nest. The effectiveness of different degrees of hunger as well as of different intensities of electric shocks was also investigated. The author states that punishment is more effective than reward and the more intense the shock or hunger, the quicker does the rat learn.

The work of Hoge and Stocking (24) upon the relative value of reward and punishment in the formation of the visual discrimination habit has recently been repeated by Warden and Aylesworth (52). Thirty white rats were used, ten in each incentive group, the groups being: reward, reward-punishment, and punishment. In the main, this work confirmed the principle set forth by the previous investigators that a combination of reward and punishment is more effective than either punishment or reward used alone. Of the last two, punishment alone seems to be more effective than reward alone.

In 1922 Johanson (28) studied the effect of punishment upon reaction time, using three adults. The stimuli were presented in three ways: (1) normal presentation in which no conflicting factors or conditions were given; (2) incentive presentation in which the subject was informed that the time he took to react would be presented to him before the next succeeding reaction occurred; (3) punishment presentation in which subject was told that he would

receive an electric shock only when he was not reacting quickly enough. To offset practice effects punishment was introduced at different times for each of the three subjects. In this presentation he was not given his reaction time. The results showed that the average improvement for the three subjects was 6.0 per cent for the incentive series, 14.8 per cent for the punishment. He concludes that "one possible explanation is that a state of much keener attention produced by incentive and by the expectant punishment, motivates reactions very much prompter than those normally obtained under quieter and more restful conditions. . . . That a state of keener attention or concentration of the attention was involved is attested to not only by the premature reactions, but also by the feeling of the subjects. Especially during the punishment set did the subjects evidence the usual signs of close attention as well as copious perspiration and nervous tension." ((28), page 51)

Rexroad (40) administered an electric shock as punishment for inaccuracy in continuous multiple choice reactions. Sixty adult subjects were used, one-half of whom received shock and one-half of whom acted as a control group. The findings show that subjects given the electric shock for inaccuracy were 15 per cent more accurate, or learned faster. The author points out that punishment has three effects: it may act as an incentive, it may have an instructive effect, or it may act as a disruptive and distractive influence.

An investigation on the effect of incentives on discrimination as measured by the Galton bar was made by Hamilton (22) who used as subjects six groups of ten undergraduate women. All groups were given a practice series of fifty trials on the first day. On the second day the subject performed a series of five trials in the same manner as the first day, after which forty-five trials under some conditions of incentive were given. Group A was given punishment, a bell ringing when a poor discrimination was made; Group B was rewarded by the ringing of a bell when discriminations were made correctly; a control Group C performed the series as given on the first day; Group D was called the guess-with-punishment group and was given the same set-up as Group A except that in addition the subject was required to guess the direction of his error. In the told-with-punishment group the situation was similar to that of Group A except that the subject was told the direction of the error that was "long" or "short." In Group F, the knowledge group, no bell

was used but the experimenter said "long," "short," or "right." Findings indicate that (1) mere continued repetition of this simple discrimination brought practically no change in accuracy; (2) the attitudes of reward and punishment created by the incentive stimulus were not reliably different in their effectiveness; (3) the subjects did not better their responses after having been given an opportunity to guess the direction of error; (4) definite knowledge in the absence of reward or punishment was not as effective as the incentives without such knowledge; (5) the emotional nature of the reward or punishment attitude seemed to heighten the attention to the task at hand.

RELATIVE EFFECTIVENESS OF KNOWLEDGE OF VERSUS NO KNOWLEDGE OF PROGRESS AND RESULTS OF PERFORMANCE

In a study on practice without knowledge of results Judd (30) states that the striking fact which appears in the results of these ten days is that practice brings little if any change, the first and last day differing from each other about as did the first and second. There is no motive for improvement. The subject cannot see his results and cannot judge of their success or unsuccess. Spencer (43) criticised Judd's statistical treatment of his results and repeated the work using four subjects. He reported that some slight improvement resulted from practice even without knowledge of results.

The purpose of Arp's (2) study was to investigate the relative effectiveness of complete versus partial awareness of results as evidenced by efficiency on the ergograph. Three subjects were given three sets of experiments, each of which were divided into four parts; in two series the results were in full view of the subjects while in the other two series they were hidden. The findings indicate that both the absolute amount and the rate of work done under conditions of knowledge of results exceed that done under conditions of ignorance of results.

Chapman and Feder (11) investigated the effect of external incentives on improvement. Three tests were given to fifty-six children from grade 5A in the Cleveland Observation School. The tests were: practice in addition (ten minutes); substitution (five minutes); and cancellation (one minute). The subjects were then divided into two comparable groups on the basis of initial scores in the addition test. The incentives used were: (1) each individual's results of the previous day were published; (2) the point reached on the last occasion

by the subject was marked in heavy blue pencil on sheets presented for the day's work; (3) the general improvement of the class was presented in the form of a graph; (4) credits were given in the form of stars to those who stood on the previous day in the upper 50 per cent of the class; and to those who, in amount of gross improvement, were in the upper 50 per cent. The general trend of the results shows that the incentives exerted a considerable effect on the amount produced except in cancellation which the authors explained as possibly being due to the fact that a one-minute test was not sufficiently long enough for the novelty element to wear off.

An experimental study of the effectiveness of incentives was made by Book and Norvell (8) in 1922. They gave four simple learning tests to 124 college students. For each test the total group was divided into two sections, one of which acted as a control for two-thirds of the practice time and the other group as a stimulus section; for the last one-third of the time this procedure was reversed. The group acting as stimulus section was given the results of their practice in each test and was urged at the beginning of each practice to study the records and try to do better. The findings show that (1) the stimulus section made more improvement with a given amount of practice than did the control groups; (2) the stimulus group, which had been making rapid and continuous improvement, ceased suddenly to show such gain when the incentive was removed; (3) the control group which had been making little or no progress in practice began suddenly to improve and continued to gain more rapidly and consistently when definitely interested in the improvement as such, when made to feel that further improvement was possible; and (4) the same groups of individuals not only ceased to gain in their scores but also became more careless about their work when not interested in their improvement as such. When they were subjected to special groups of incentive they began at once to increase their output and accuracy.

Kirby (32) investigated the influence of knowledge of results upon practice in addition and division. He used 1350 New York public school children from the third and fourth year classes. After the practice period the children were encouraged to tell about the number of problems which they had worked. In addition, at the beginning of the next practice period, the exact score of each child in the preceding day's work was read, both the number of columns worked and the number correct. Results were in favor of motivation. "At

the end of the test there were frequent expressions of regret that the time was up, but almost no indications of indifference, of ennui. . . . Much of the gain in efficiency must be due to the children's knowing their previous success, which in turn induced a pleasurable emotional tone to inspire continued intense effort and to strengthen associations formed." ((32), page 43)

Two experiments are reported by Thorndike ((51), page 430) which he considers "a crucial test of whether the after-effects of a connection do in fact strengthen or weaken it." In experiment A, five subjects were required to estimate in centimeters the length of fifty strips of paper, using as a standard a ten centimeter strip placed before them. A control series of fifty judgments were made, no help being given the subject. A series of seven or eight training periods were given during which the experimenter told the subject whether his judgment was "right" or "wrong." Following this a final test period was given in which nothing was said to the subject. A comparison of the first and final tests showed that the average percentage of reduction was 61 per cent. Seven subjects who had the same amount of training but were not given knowledge of successes or failures made an average reduction of 7 per cent. In experiment B a similar procedure was carried out except that twenty-four subjects were blind folded and asked to draw a three, four or five inch line in response to the appropriate request of the experimenter. The average gain in percentage of correct lines on the last test over the first was 12 per cent. Of six subjects who had no knowledge of their successes or failures three improved and three made more errors.

The purpose of Ross' (41) study was to inquire into the relation between knowledge of progress and achievement in one form of motor learning. The subjects were fifty-nine students in psychology. Three sections of equal size and of substantially equal achievement were formed on the basis of scores made in the first practice period. Section I was given full information as to their scores and progress day by day, Section II was given only partial information, and Section III was given no information at all. All sections were urged to do their best. On the eleventh day the procedure was reversed. The author states in conclusions that "even in highly motivated groups, as undoubtedly these were, the addition of a single other motivating factor, namely knowledge of results, is sufficient to give the pupils with such knowledge a distinct su-

periority over the others, and the degree of superiority is roughly proportional to the amount of information possessed. To what extent this result is to be explained by the hypothesis that additional sources of energy were released in subjects who knew their progress due to the mild emotional effects accompanying rivalry, either social rivalry or self-rivalry, is not known, but this was probably a factor of considerable importance." ((41), pages 345-346)

INFLUENCE OF PRAISE AND REPROOF UPON PERFORMANCE

Gates and Rissland (18) have studied the effect of encouragement and discouragement upon performance using as subjects seventy-four Barnard college students, who were given (1) the three-hole test and (2) the color-naming test. After a practice series every first subject was given encouragement, every second subject discouragement, and every third subject was merely asked to repeat the experiment. "The results of this study seem to show then, a very slight difference in average improvement or even in percentage of individuals who improve in the three groups. In this, the facts found are similar to those observed in experiments on fatigue, lack of fresh air, sleep, or food; the external factors seem to be of relatively little importance in determining the score. Such difference as there is seems to be in favor of encouragement or discouragement rather than mere repetition. We might say then, ... that it is a little better to make an encouraging than a discouraging remark; that relatively poor individuals are more likely to be unfavorably affected by discouragement than are relatively proficient persons; that the effect of these incentives does not seem to be constant for the two tests." ((18), pages 23-24)

One of the most important studies on motivation is that of Hurlock (26) on the value of praise and reproof as incentives for children. The author gave the National Intelligence Test to a group of 408 children selected from the third, fifth, and eighth grades of a New York City school. On the basis of an initial test, age, sex, and race, three equivalent groups were formed: one group was praised, one reproofed, and the third acted as a control. A second test was then given. The author found that both praise and reproof were approximately of equal value. There were age differences, the older children responding more to both praise and reproof than the younger children. Boys did better following praise and reproof than did girls. The boys responded better following reproof and the girls responded better following praise. As to mental age dif-

ferences some incentive was more essential for "superior" children than for "inferior" children. Of the two types praise was somewhat better for "inferior" children and reproof for "superior" children.

Briggs (9) reports a study carried out with high school students, the purpose of which was to obtain objective evidence as to the relative merits of praise and censure. Two teachers who were accustomed to use severity with their pupils were asked to conduct the experiment. First, they made new type tests on three lessons, two-fifths on judgment and the remainder on factual memory. The first test was given by each teacher to his four classes. After glancing over the papers, Teacher A reprimanded his pupils severely, threatening them with punishment if they did not do better work; Teacher B commended his pupils and encouraged them to do even better on the next lesson. The following day the second test was given, and after looking over the papers the teachers reversed their comments. On the third day the last test was given. It was found that 87 per cent of the pupils made better scores after commendation and encouragement.

An investigation was made by Gilchrist (19) to determine the extent to which praise and reproof affect a pupil's work. The subjects were fifty members of a class in educational psychology, who were given the Courtis English Test, Form 4B. The group was first given an initial test and then divided into two groups without regard as to how well the individual members of the class had done in the test. After a lapse of time, the length of which is not stated, the two groups were given the same test again. This time Group A was reproved and Group B was praised. The results showed that Group A (reproved) made no improvement in spite of practice. The effect of praise and reproof was indicated by a change in efficiency of 79 per cent. The members of the class who had done well in the first test did not do as well in the second test. In so far as Courtis English Test, 4B could be considered a test of mentality, the "good" students made the higher scores in the first test; the "poor" students, on the other hand, made lower scores. The results of the second test then show the "good" students to have been more influenced by comment than were the "poor" students. It may have been that the former were unaccustomed to disparaging remarks, whereas the latter were inured to them.

The effect of incentives upon the constancy of the I. Q. has been

studied by Hurlock (27), her purpose being to determine how much influence an encouraging or discouraging comment preceding the test would have upon the test results. The subjects were 273 boys and girls from two New York City schools. The National Group Intelligence Tests, scale B, Forms 1 and 2 were used. All children were given an initial test and then formed into three experimental groups on the basis of I.Q. scores obtained in the initial test, chronological age, and sex. One group acted as a control, one was praised, and the other reproved. The results show that incentives in the form of praise and reproof raise the average I.Q. scores seven points while less than one point increase resulted from practice alone.

A study on the influence of commendation or favorable response by the tester has been reported by Nelson (39). The subjects tested were preschool children. Research assistants, concealed from view of the child, recorded the instances where praise or commendation were given by the experimenter during the course of a routine intelligence examination. The author states that "the one significant result of this study was that it revealed the routine use of expressions of commendation. The data indicated that in no psychological examination were there favorable responses by the examiner to the test situation in less than 60 per cent of the time, and that the use of such expressions tend to be much more frequent. The median falls at 84 per cent with a range of from 63 per cent to 97 per cent. For this group there was a tendency for the younger children to receive more praise on the average than the older children. Mental age as well as chronological age would seem to be a determining factor since the children with lesser degree of intelligence received the most praise. The data are meager, but subjective evaluation of one's technique would seem to support the contention that in difficult situations there is a tendency to give more explicit and enthusiastically toned praise, probably as a method of encouragement. There is a consistent tendency to use favorable responses as a means of smoothing the testing situation, easing a child through difficulty. In one-third of the total number of cases given three successive trials on the Seguin Test, the amount of praise given on each of the three trials definitely corresponded with the degree of success on the trial, as measured by the time and error scores."

EFFECT OF PURPOSE AND AIM

Wright (55) made an investigation of some effects of incentives on work and fatigue. Using the Cattell ergograph, four subjects

were required (1) to work as hard and as long as they could but with no idea of securing any specified result, and (2) to do a specified task as long as their strength endured. Under (1) the subject was not allowed to watch his progress; under (2) he was not only allowed to watch progress but in addition to this, part of the time was stimulated to action by being told to count his strokes. Blocks were placed under the carriage and the subject was requested to see how many times he could reach the block. The results of all four subjects showed a gain in work done when incentives were present. The subject accomplished more work when working under the mental stimulus of having a set task to perform than he did when working without a definite aim. This the author interprets as being due to the mental attitude of the subjects under the different conditions of the experiment. The fatigue accompanying the work was not so great when the subject was working under the stimulus of a definite aim, notwithstanding the fact that at the same time he produced an increase in the amount of his work.

Fluctuations in mental production when motivation is the main variable were investigated by Knight and Remmers (33). The Thorndike addition sheets were used and given to an experimental group of ten college freshmen who were undergoing a severe probation period preparatory to fraternity membership, and to a control group consisting of fifty-four juniors, mostly women. A member of the fraternity told the freshmen in the experimental group that they were to assist in a scientific experiment, and, since scores on the test were involved in final election, they were to push themselves to the limit. The results were in favor of the motivated group. "The difference between a genuine motivation such as the freshman had, and the kind of motivation a college class contains, is a difference which not only offsets extreme fatigue but further offsets freshman vs. junior ability, and in addition produces twice as much work per unit of time with equal accuracy." ((33), page 223)

The purpose of Moore's (37) experiment in opening a glass box to get toys was to measure the effort of young children in working toward a deferred goal. The effect of the presence or absence of toys within the box was investigated and it is stated by the author that it was apparent that most of the children who were given this series of repetitions did not make as great an effort when the toys were removed. When working at the empty box, there was more restless-

ness and many more extraneous, uncoordinated movements. In several cases the effort was renewed when the toys were returned.

INFLUENCE OF ATTITUDE UPON TEST PERFORMANCE

Thorndike ((48), pages 24-25) states: "It is a general law of behavior that the response to any external situation is dependent upon the condition of the man as well as upon the nature of the situation. . . . Only a little less obvious should be the fact that the attitude or set of the person decides not only what he will do and think, but also what he will be satisfied and annoyed by Purposive behavior is the most important case of the influence of attitude."

Ruger (42) and Book (8) in experiments on learning process give the introspective accounts of various subjects as to the influence of attitude. Ruger (42) states that the attitude of confidence was a great aid in the successful solving of puzzles. Book (8) speaks of the influence of a favorable "feeling tone." The mental set with which a person approaches a task is a large factor in success or failure. Bronner (10) found in the mental testing of court cases that attitudes of resentfulness, feeling of shame, embarrassment, and general nervous excitement markedly influenced the results and it was only after the causes for these attitudes had been removed that a truer test of ability was gained.

Dunlap and Snyder (14) studied the effect of practice on I. Q. scores. The subjects were forty-four college men who were tested four different times with four different forms of the Army Alpha Test. Three weeks elapsed between each test. The results showed general improvement on the second and third testing but a general falling off on the fourth. The main reason given for this deterioration was that students reported that they were bored at having to take it over again.

The purpose of Sullivan's (44) study was to ascertain: (1) What changes take place in the process of learning by the individual because of the presence of a given attitude with the learning response? (2) Are different attitudes equally effective in varying the learning situation? The subjects were a group of adults of superior and inferior intelligence and a group of school children. The task given the adults was that of learning nonsense syllables. The children were given the problem set by the Miles pursuit pendulum. The data show that the time required to learn a memory series is increased by knowledge of failure and decreased by knowledge of suc-

cess; women are more affected than men; subjects for whom the task is difficult are more affected than subjects for whom it is easy; subjects of high intellectual capacity are more affected by knowledge of failure, subjects of low capacity by knowledge of success; certain adaptations are made by adult subjects in the case of continued failure report.

Bird (4) had as the aim of her study the determination of one of the causes of variety of achievement in the learning of one hundred healthy young children of normal intelligence. Children four to six years of age were given one of the stages of the Henry Barnard School method of mastering reading. The author states that of the group of one hundred, thirty were found with habitual personality handicaps that interfered with learning, while thirty-seven showed unmistakable but slighter affective disturbances. Causes given were attitudes of shyness, lack of self-confidence, fear of the task. Eight children were definitely delayed in their progress whenever praise was not forthcoming. Sometimes it was sought as a means of encouragement to further effort, but oftener to satisfy the ego. In the last analysis the same affective disturbances handicapping these children function largely in the failure of adults to achieve to the extent of their mental ability. The over-timid, the anti-social, the chronic introvert, the braggart, the annoying egotist, the excessively sociable are all victims of habits that divert attention from the task in hand.

The importance of keeping the child encouraged has been emphasized by Terman ((45), page 125) who points out that "Nothing contributes more to a satisfactory *rapport* than praise of the child's efforts. Under no circumstances should the examiner permit himself to show displeasure at a response, however absurd it may be. In general, the poorer the response, the better satisfied one should appear to be with it. An error is always to be passed by without comment, unless it is painfully evident to the child himself, in which case examiner will do well to make some excuse for it. . . . Exclamations like 'fine!' 'splendid!' etc. should be used lavishly. Almost any innocent deception is permissible which keeps the child interested, confident, and at his best level of effort."

Alpert (1) in a study of problem-solving in young children includes a section on the factors preventing arousal of insight and its natural consummation in a practical success. She lists such attitudes

as (1) self-consciousness, (2) lack of confidence, (3) lack of interest, (4) discouragement.

INFLUENCE OF SUCCESS AND FAILURE UPON PERFORMANCE

In a study on the motor skill of preschool children Goodenough and Brian (20) include a section on the effect of success upon the type of performance. When the nature of the errors was examined, it was found that the greater number of children showed some tendency to throw beyond the mark. This tendency, however, was not distributed at random, but was far more evident in the errors immediately following successful throws than in those following unsuccessful throws. This rather curious point calls to mind Münsterberg's observations on the expansive outward-going movements accompanying the pleasurable emotions. The extent of the tendency may be judged from the fact that of twenty subjects, seventeen showed this characteristic to some extent, while in the three remaining cases the differences were negligible. For the entire group of twenty cases, the difference between the proportion of overthrown errors following successes as compared to those not following success is 5.47 times the standard error. Analysis of individual learning curves by means of study of daily record sheets appears to throw considerable light upon a number of factors influencing success. Among these may be mentioned emotional factors, including not only changes in motivation but also over-confidence or undue caution.

Hicks ((23), page 51) in a study on the acquisition of motor skill in young children also includes a section on the relation of success to other trials. He concludes that "there is evidence then that the ringing of a bell was related to the throws which immediately preceded and immediately followed it. The throw in which the bell was rung was probably affected by the practice preceding it and most by the practice on the throw immediately preceding. In like manner, the throw in which a bell was rung probably had its greatest carry-over to the throw immediately following. This may have been due to the factor of motivation which would probably function best immediately after throws which either hit the bell or came very close to it."

SUMMARY OF THE CHARACTERISTIC FINDINGS IN THE FIELD

The findings will be discussed under (1) aims and purpose, (2)

materials and apparatus used, (3) subjects, (4) methods, and (5) results.

The aims and purpose of the studies of the influence of incentives upon performance have been concerned with the relative effectiveness of reward, punishment, praise, commendation, encouragement, approval, reproof, discouragement, disapproval, ridicule, aim, purpose, no aim, no purpose, full or partial knowledge of progress or results, complete or partial lack of knowledge of progress or results, success, failure, and attitude as it affects performance. (This report has not included a review of literature in drives, competition, rivalry, or social facilitation.)

The materials and apparatus which have been used to measure the effectiveness of these various incentives have consisted of both physical and mental tasks. These include learning nonsense syllables, addition, subtraction, division, multiple-choice, discrimination, learning problem set by Miles pursuit-meter, ergograph, intelligence and achievement tests, learning to run a maze, reaction time.

The subjects have been both animals and human beings. The latter include adults as well as individuals at the grade school, high school, and college levels.

The methods have been experimental for the most part. It may be pointed out that in many cases there was no control group or other means of taking into account practice, fatigue, or carry-over effects. In some investigations the subjects were tested individually, in others they were tested in a group even though the main purpose was not that of studying the effects of rivalry.

Because of the varied aims and purpose, the differences in age, range, and kind of subjects used, the wide differences in material and apparatus, and especially the differences in method, it is difficult to compare adequately the results of these investigations. These differences should be kept in mind when evaluating the results.

1. When reward, punishment, or a combination of reward and punishment are used with animals it seems evident that the combination of reward and punishment is more effective than either reward or punishment used alone. Of these latter two when used alone, punishment is more effective than reward.

Of the three studies where reward and punishment were used with humans, two indicate that punishment is more effective than reward while in one no reliable differences in their effectiveness were found. In all three cases there were indications that the attention was heightened during the motivated series, particularly in the case of punishment.

2. Of the eight studies dealing with the relative effectiveness of knowledge versus no knowledge of results reported here, the findings of all but one indicate that either partial or complete knowledge of progress or results brings with it a decided improvement in performance when compared with a group who were not given such knowledge. Superiority may be said to be roughly proportional to the amount of information possessed.

3. Five experimental studies dealing with the influence of reproof and praise are summarized. Of this number three resulted in findings which showed a fairly decided percentage of improvement in favor of praise, one found that praise and reproof were superior to repetition but that the two former were approximately equal in value, while one study found very little improvement at all but that such differences as there were seemed to favor encouragement. There seem to be sex, mental age, and chronological age differences.

4. The three studies on the effect of purpose and aim reported, indicate that a purpose or aim facilitates learning or results in improvement in performance over individuals working with no purpose or aim.

5. The influence of "mental set," *aufgabe*, attitude is pointed out by various writers. Because of it the situation is supposedly undertaken with a more or less definite emotional tone, which may or may not be favorable for the effective performance of the task. Reports indicate that attitudes of self-confidence, shyness, success, failure, etc., all affect performance.

6. Two studies have pointed out that there is a tendency for greater release of energy following successes when throwing at a target in that errors immediately following a success are usually beyond the mark.

PROBLEMS IN MOTIVATION OF YOUNG CHILDREN

Since no extensive or systematic work on the influence of incentives has been done using children under eight years of age, we find that there is a need of obtaining data on the entire range of problems which have been investigated with adults and older children. Accordingly, we may ask the following questions concerning the influence of incentives as used with young children.

1. Is the use of some external incentive in learning, habit formation, or as a means of social control more effective than no such incentive?

2. What are the relative values of different incentives as repetition, praise, commendation, encouragement, approval, success, reward vs. reproof, discouragement, disapproval, failure, or punishment; complete or partial knowledge of results vs. complete or partial lack of knowledge of progress and results; working with a purpose or aim vs. having no purpose or aim, etc.?

3. What are the relative values of different intensities of any of these incentives?

4. What is the effect of the continued application of any one of these incentives?

5. What proportion of positive and negative incentives will insure maintenance of a certain level of efficiency?
6. Are there chronological age, mental age, or sex differences?
7. What is the relative value of giving these incentives privately, before another adult, before another child, or before more than one child or adult?
8. At what age do rivalry and competition begin?
9. Under continued application of any of these incentives does one find defense or expansive mechanisms beginning to operate? If so, when and how can these mechanisms be reconditioned or eliminated and which types of incentives are most effective in so doing?
10. What is the relative effectiveness of any one of these incentives as given in one task upon the performance of another task? If there is a facilitating or inhibiting after effect, how long is it or can it be maintained?
11. What is the relative effectiveness of different combinations of incentive?
12. What influence does incentive have upon initiative, persistence, and self-confidence?

CHAPTER III

STATEMENT OF THE PROBLEM, METHOD, AND SUBJECTS OF THE INVESTIGATION

STATEMENT OF THE PROBLEM

Up to this point we have considered the importance of motivation in the life of the child and have reviewed the evidence that different types of motivation influence the responses of the individual as he performs a task. From this latter survey we have summarized the characteristic problems and results in the field and have attempted to show some of the problems of motivation which concern the young child. These problems are, of course, too numerous to include in one study. The present investigation represents, therefore, but one approach to certain aspects of the problem of motivation and its effectiveness for the performance of a task. The emphasis has been upon the external types of incentives rather than upon the more internal types of motivation designated as drives, urges, needs, and wants.

It has been the purpose of this investigation to obtain data on the relative effectiveness of a number of different types of external incentives which have been given under controlled and experimental conditions with four comparable groups of children ranging in age from two to eight years. The following types of incentives have been used: control-motivation, knowledge of progress and results of performance, no knowledge of progress and results of performance, success-repetition, failure-repetition, success-praise, failure-reproof, success-reward, and failure-punishment.

The main specific objectives of this study have been stated in the form of the following questions:

1. Is some motivation more effective than control-motivation for the performance of a task?
2. Is some motivation in addition to success more effective than success-repetition?
3. Is some motivation in addition to failure more effective than failure-repetition?
4. What is the relative effectiveness of success-praise versus success-reward?

5. What is the relative effectiveness of failure-reproof versus failure-punishment?
6. What is the relative effectiveness of success-repetition versus failure-repetition?
7. What is the relative effectiveness of success-praise versus failure-reproof?
8. What is the relative effectiveness of success-reward versus failure-punishment?
9. What is the effect of each of the above conditions of motivation upon the performance of another task?

METHOD

In order to be able to compare the relative effectiveness of any one of these incentives with any other it was necessary to use equivalent groups, since to give so many variables to each subject would lead to confusing results because of practice, fatigue, and carry-over effects. It seemed necessary to have at least four groups in order to include some data on each of the incentives. If possible, it would be necessary to have at least twenty-five to thirty subjects in each group if results were to be adequately analyzed using statistical measures. The groups were to be equated on the basis of the initial score on the first task.

The general method was, therefore, experimental, using four equivalent groups which were formed on the basis of initial scores on the Motivation-Dynamometer Test. Each subject was tested three times, the interval between each testing day being one week apart. The total range of time elapsing between the first and last test for each child was limited to three weeks in order that maturation or increase in strength itself should not exert a constant effect on the results. Each time the subject was tested he was given two tasks. The first task, which was based on the principle of the hand dynamometer, was devised by the investigator and is described in Chapter IV. It was immediately followed by another task which also involved motor ability, namely, the Perforation Test which is described by Baldwin and Stecher.⁸ In addition to the quantitative results gained from numerical scores on performance in these two tasks, certain qualitative data were obtained. A recorder checked the presence, or degree of presence, of certain qualitative responses or attitudes which the subject made or exhibited during his performance of the tasks. The total time taken for testing each subject on

⁸ Baldwin, Bird T., and Stecher, Lorle I.: *The Psychology of the Preschool Child*. New York: D. Appleton, 1924. Pp. 305. (p. 95-98)

any one testing day was approximately ten minutes. At this rate as many as thirteen to fifteen children could be tested in a morning session which convened from nine until eleven-thirty o'clock, and as many as ten to thirteen children during an afternoon session.

FORMING COMPARABLE GROUPS

For various reasons it was not possible to use a common procedure in equating the four groups, namely, that of giving the initial test and then matching the groups on the basis of scores in that test before any other variables were applied. This was, in part, due to the fact that the total number of subjects needed did not at the time seem to be available in Iowa City. It was decided, therefore, that after the initial test the subjects should be assigned to groups on a purely random basis. Each of the separate schools and groups of children tested were fairly homogeneous as to age range included in each group. For instance, Group I at the preschool laboratories of the Iowa Child Welfare Research Station included only children between the ages of two years to two years, eleven months. Similarly the Home Laboratory enrolled three-year-old, Group II four-year-old, and Junior Primary five-year-old children. It was desired to have subjects of each age level in each of the four experimental groups.

Accordingly, as each of these large groups was tested the names of the subjects were listed in alphabetical order and then numbered. A pack of cards was similarly numbered, the pack then shuffled and dealt into four piles constituting Group A, Group B, Group C, and Group D. Each subject was then automatically placed in the group in which his number fell. It was thought that by using this random method of placing a child in an experimental group that the means and standard deviations of each of the four groups on this initial test would be fairly comparable if a sufficiently large sample were obtained. As will be shown under the section treating the results, this method proved to be an adequate one as it was necessary to eliminate only two cases in order to make the means and standard deviations of each of the four groups comparable within one probable error.

PLAN OF TESTING

With the above method of grouping in mind, we may now proceed to a consideration of the general plan of the testing program. Each subject was tested three different times, approximately one week

apart. For the most part the groups were located in different buildings. As the apparatus was not conveniently portable the experimenter usually spent one or two days of the week with a group, then moved on to another group. No more than four different school groups were tested during any one week. If a child were absent on the day that the experimenter was testing children in his group, that subject was dropped. However, in several instances where more than one group was housed in the same building or when the experimenter stayed more than one day in the same group a subject was tested after his return to school, provided the interval did not exceed more than one week.

Since each subject was tested three different times, approximately one week apart, the general method of procedure according to the first, second, and third weeks will be considered. The Motivation-Dynamometer Test representing the first task was given to the Initial Group and Groups A, B, C, and D in Series I, II, and III with seven trials each, followed by the Perforation Test, or the second task, with three trials each. This plan is presented in the following concise form:

Group	Presentation	Condition	Group	Trials
Task 1—Motivation-Dynamometer Test				
First Week—Series I				
Initial	1	Control-Motivation	Preliminary	7
Second Week—Series II				
A	1	Control-Motivation	Control 1	7
B	2	Motivation	Success-Repetition	7
C	2	Motivation	Success-Praise	7
D	2	Motivation	Success-Reward	7
Third Week—Series III				
A	1	Control-Motivation	Control 2	7
B	3	Motivation	Failure-Repetition	7
C	3	Motivation	Failure-Reproof	7
D	3	Motivation	Failure-Punishment	7
Task 2—Perforation Test				
Series I, II, III				
Initial Group and Groups A, B, C, D				3

First Week

During the first week all subjects were given the control preliminary series of seven trials on the Motivation-Dynamometer task

under Presentation 1, each one minute in length, followed by three trials on the Perforation Test.

Second Week

One week later the subject was seen again. In the meantime the above mentioned method of placing a child in an experimental group had been worked out so that when the subject came in for his second test he was given the type of motivation required by the demands of his group. Group A acted as the control group and received the same type of motivation as was given to all subjects during the initial or control preliminary series in the first week. Groups B, C, and D were all given success under Presentation 2. In addition, Group B was given only repetition of directions between trials, Group C was given praise between trials, and Group D was given reward.

Third Week

This time, as before, Group A was given Presentation 1 and was known as Control 2. Groups B, C, and D were given failure on Presentation 3. In addition, Group B received repetition of directions between trials, Group C was given reproof, and Group D was given punishment.

Since all four groups were comparable to start with, and because a control group was used in addition, we are able to compare not only the relative effectiveness of incentives as given in any one week (praise with reward), but also the effectiveness of incentives between weeks (as success with failure).

SUBJECTS

In working out the procedure, a preliminary group of twenty-one children enrolled during the summer session in Group II of the preschool laboratories of the Iowa Child Welfare Research Station was used as subjects. No children in this group were used in the main experimental group. The range in chronological age was from three years, ten months to five years, five months with an average age of four years, four months.

There were 259 children in the main experimental group who were given at least one test of the three test presentations. These subjects were secured from four sources: seventy children from the preschool laboratories of the Iowa Child Welfare Research Station; forty-one children from Grade I and Grade II of the University

Elementary School, State University of Iowa; seventeen children from the Nursery School at Iowa State College, Ames, Iowa; ninety-three children from four kindergartens of the public school system, Ames, Iowa. The age range of the entire group was from two years, three months to eight years, with all ages taken from the date of the first presentation of the test series. The distribution of the subjects according to age in months was:

Age, Months	Children
24 to 35	9
36 to 47	20
48 to 59	28
60 to 71	86
72 to 83	48
84 to 96	20

Two years of age refers to those children who range in chronological age from two years, zero months to two years, eleven months and similarly for all other ages. There were, then, nine two-year-olds, twenty three-year-olds, twenty-eight four-year-olds, eighty-six five-year-olds, forty-eight six-year-olds, and twenty seven-year-olds.

The children came from homes of varied occupational levels and economic status. In general, those subjects who attended the pre-school laboratories of the Iowa Child Welfare Research Station, the nursery school at Iowa State College, and two of the kindergartens of the public school system came from homes of the professional class. The children from one kindergarten came from homes of the average class and those in the remaining kindergarten group came from homes below the average.

CHAPTER IV

APPARATUS, MATERIALS, AND TEST PROCEDURE USED IN THE STUDY

In this section the apparatus, materials, and test procedure which were used to obtain a measure of the effect of the various types of motivation upon performance are considered.

Two tasks were presented the subject each time he was tested. The first task was designated the Motivation-Dynamometer Test and was devised by the investigator; the second task was known as the Perforation Test and has been used at the Iowa Child Welfare Research Station for a number of years.

MOTIVATION-DYNAMOMETER TEST

There were four essential considerations to be kept in mind in devising the task which was to be used in measuring the effectiveness of the various types of incentives under investigation: (1) the task must be such that the units of measurement were objective and quantitative, whether they should be in time, amount, or accuracy units; (2) the task should be one which could be used throughout a wide age range, namely, from two to eight years; (3) the task must be one in which success and failure could be insured and yet at the same time allow the experimenter to measure the subject's actual performance; (4) it should allow for visual motivation. It follows from the second consideration that the task should be fairly interesting to all ages concerned. Again, it should also be adjustable in difficulty levels for the entire age range in question in order that it should not be too easy for the older child nor too difficult for the younger subject.

These requirements made it almost impossible to use a task which depended upon mental ability alone since it would be difficult to insure success and failure over such a wide age range. It was, therefore, necessary to fall back upon a task involving physical capacity. A survey revealed no test in use which met the general requirements already indicated. It was, therefore, necessary to devise a piece of apparatus which would more nearly fulfill these needs. The writer is indebted to Dr. George D. Stoddard, director of

the Iowa Child Welfare Research Station, for the suggestion of the basic principle which was finally used, that of squeezing a rubber bulb and thereby forcing the liquid in the bulb up a glass tube connected with it. After working with the problem for a short time it was apparent that a measure of the amount of energy or effort expended through the muscles of hand and arm was being attempted. It was somewhat similar in principle to the Smedley dynamometer which measures strength of grip.

Buford Johnson (29) who has used the Smedley dynamometer with young children states, "The breadth and thickness of the bars prevent the smaller children from getting a firm hold and probably causes a discomfort which handicaps the performance in succeeding trials."⁹ Another criticism of the instrument was that its tension is adjusted to the adult hand. This is not altogether suited to the young child's muscles. Experiments in physiology have pointed out that the amount of work accomplished by the muscle is greatest when the load is about one-half of what the muscle can do, and least when it is either too light or too heavy. A third criticism of the instrument for use in this study was that when used it was held at the subject's side and thus did not allow for visual motivation.

There were, then, four requirements specific to the apparatus which should be kept in mind: (1) it should be adjusted to the size and shape of the young child's hand; (2) it should avoid, if possible, causing the subject discomfort or pain in the manipulation of it; (3) the tension adjustment should be more nearly suited to the young child's capacity; and (4) it should allow for visual motivation.

The apparatus which was devised to meet these requirements was composed of three units designated as: (1) the main system; (2) the adjustable chair and arm rest; and (3) the train system.

MAIN SYSTEM

This unit consisted of the following systems: (1) the basic system; (2) the system which maintained liquid stationary at its highest point in the tube; (3) the system which permitted liquid to be lowered; (4) the lever system; and (5) the system by which success and failure could be controlled.

⁹ Johnson, Buford J.: *Mental Growth of Children in Relation to Rate of Growth in Bodily Development*. New York: E. P. Dutton, 1925. Pp. xix, 160. (p. 40-41)

Basic System

Figure 1 shows the general arrangement of all systems of the main system. The essential details of the basic system, however, may be seen in Figures 1 and 2. A rubber bulb holding about eight

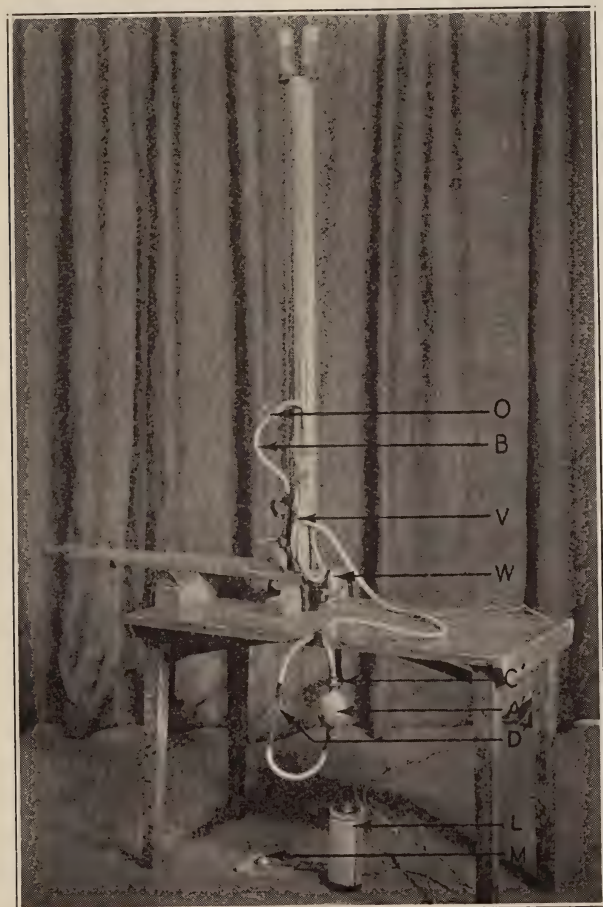


Figure 1 Motivation-Dynamometer Test, Main System, Unit I:
All systems

- O Stopcock
- B Rubber tubing
- V Joining of system worked by experimenter with main system
- W Liquid forced into main system at V through W
- C¹ Valve
- A¹ Rubber bulb
- D¹ Stopcock
- L Dry cell battery
- M Push button

cubic centimeters of liquid (water plus vegetable coloring) was placed in a depression made in the baseboard and was allowed to rest vertically against the upright. This bulb was of the ordinary

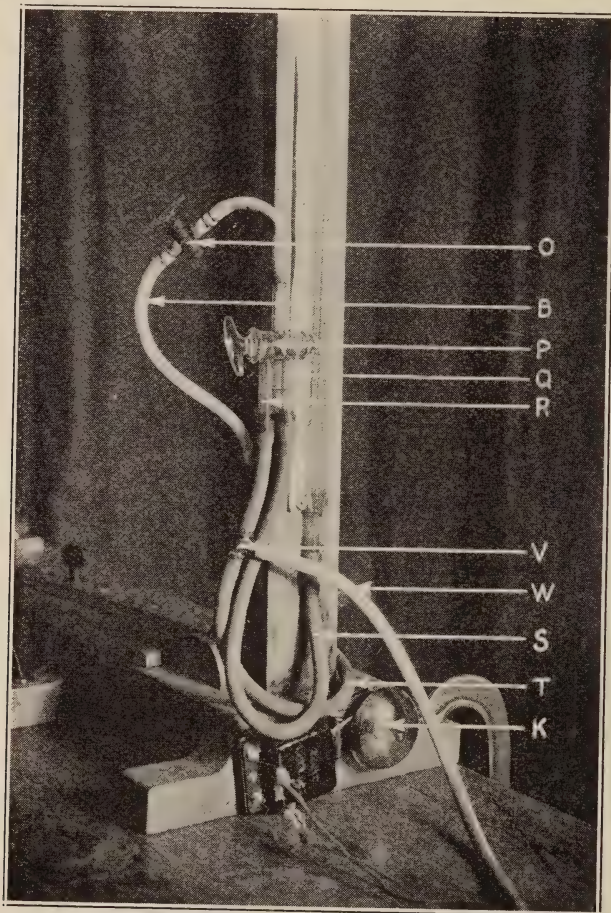


Figure 2 Motivation-Dynamometer Test, Main System, Back View

- O Stopcock
- B Rubber tubing
- P Stopcock
- Q Outlet
- R Outlet
- V Joining of system worked by experimenter with main system
- W Liquid forced into main system at V through W
- S Back tube
- T Liquid forced into main system at V and around into the front by means of this rubber tubing
- K Bell

variety, such as might be used in a chemistry or physics experiment, except that the valves in either end had been removed and a more suitable valve inserted. This bulb was connected by means of rubber tubing and two stopcocks to either of two glass tubes which had inside diameters of 4 mm. and were 1 m. in length. One of the tubes was fastened to the back of the upright and the other was fastened to the front of this upright. When pressure was brought to bear upon the rubber bulb, the liquid was forced out of it and into either of the two glass tubes. Measurement of the amount of force expended was obtained in terms of the height of the column of liquid in the glass tube. For this purpose a paper scale, reading in metric units with zero at the bottom, was fastened to the back of the upright and to the left of the glass tube. Readings were taken in centimeters and to the nearest millimeter.

The height of the column of water in the tube was proportional to the amount of energy expended since according to the law of work and energy, as stated by Merchant and Chant (13), "work denotes the quantity obtained when we multiply the force by the distance in the direction of the force through which it acts," and because according to the law of the lever, "the force obtained divided by the force applied is equal to the inverse ratio of the lengths of the lever arms." However, it was evident that when the fulcrum was moved from one hole to another the readings in their raw form were not comparable to readings taken on any other hole since the ratio between the lever arms had been changed. In order to be able to make all readings on all holes comparable, it was necessary to calibrate the apparatus in the following way: A spring balance scale,¹⁰ which read in kilogram units, hooked over the two bars at the same position at which the child's hand was placed during a squeeze. The scales were then used to pull the movable bar toward the stationary one in the same manner as did the hand in an actual squeeze. This movement of course caused the opposite end of the movable bar to force the liquid out of the bulb and up the tube. It was desired to have a corresponding kilogram value on all eleven holes for each of the following centimeter distances up the tube: 10, 20, 30, 40, 50, 60, 70, and 80 centimeter distances. It was, therefore, necessary to have someone stationed at the back of the upright to warn the person manipulating the scales that the liquid was approaching a certain

¹⁰ The spring balance scale was obtained from the Central Scientific Co., 460 East Ohio St., Chicago, Ill.

point in the tube and that at the next signal the corresponding reading should be taken in kilogram units from the scales. The apparatus was calibrated in this way at the end of each week. Readings were taken at least twice in order to insure greater accuracy and as a check in the reporting of the values. At the end of each three weeks' testing, which completed one group on the entire three test series, the apparatus was taken apart, cleaned, and a new rubber bulb was substituted for the old one.

After having secured the paired values for each 10 cm. distance on each hole, the six readings taken during any three-weeks period in which the apparatus was used intact were averaged and the resulting values substituted in the formula, $y = a + bx$, in order to obtain the numerical expression of the constants a and b .¹¹

The following kilogram values for corresponding centimeter distances on the fifth hole were taken during the three testing weeks in February:

Centimeter Distance	First Week				Second Week				Third Week				Mean of Means
	First Reading	Second Reading	Mean		First Reading	Second Reading	Mean		First Reading	Second Reading	Mean		
10	2.9	2.8	2.85		2.9	2.9	2.90		2.8	2.9	2.85		2.87
20	4.3	4.4	4.35		4.4	4.3	4.35		4.3	4.3	4.30		4.33
30	5.7	5.5	5.40		5.3	5.3	5.35		5.4	5.4	5.40		5.35
40	5.7	5.9	5.80		5.9	5.8	5.85		5.8	5.7	5.75		5.80
50	6.7	6.5	6.60		6.8	6.6	6.70		6.6	6.7	6.65		6.65
60	7.4	8.6	7.50		7.5	7.6	7.55		7.5	7.5	7.50		7.52
70	8.4	8.6	8.50		8.5	8.5	8.50		8.4	8.7	8.55		8.52
80	9.0	9.2	9.10		9.2	9.1	9.15		9.0	9.1	9.05		9.10

The values in the last column under the mean of the means for each 10 cm. distance were substituted for x in the formula, $y = a + bx$, as follows: Let y equal centimeter distance and x the corresponding kilogram value. Then

$$10 = a + 2.87b$$

$$20 = a + 4.33b$$

$$30 = a + 5.38b$$

$$40 = a + 5.80b$$

$$50 = a + 6.65b$$

$$60 = a + 7.52b$$

$$70 = a + 8.52b$$

$$80 = a + 9.10b$$

¹¹ The original readings, averages, and formulas obtained are to be found in the manuscript copy of the monograph on file at the State University of Iowa Library, Iowa City, Iowa.

Totaling (1) $100 = 4a + 18.38b$ and (2) $260 = 4a + 31.79b$
 Then subtracting (2) from (1) algebraically: $-160 = -13.41b$
 Solving for b: $13.41b = 160$ and (3) $b = 11.94$
 Substituting obtained value for b in equation (1):

$$\begin{aligned} 100 &= 4a + 18.38 \times 11.94 \\ -4a &= -100 + 219.467 \\ (4) + a &= -29.867 \end{aligned}$$

Substituting values found in (3) and (4) we have:

$$y = -29.867 + 11.94x$$

It was, therefore, possible by use of these formulae to obtain any corresponding value of x when y was known. For example, when y equals 20, x equals 4.16; y equals 50, x equals 6.59; y equals 80, x equals 9.18. All of these calculated values are very close to obtained values. The equation, therefore, seems to fit the data. To facilitate the translation of these scores the graphic representation of these formulae was plotted and readings were taken directly from the graph. This same procedure was used for each hole.

Because the apparatus was constructed so crudely there were a great many places where friction might operate. One of these points is a certain load which must be overcome in getting started. Thus in taking all readings it was thought advisable to keep them all above 10 cm. in order to eliminate, in some measure, this factor of "load" in getting started. Therefore, if on a certain hole the subject's reading in the first trial was less than 10 cm. the fulcrum was moved to a more suitable hole in order to make the task somewhat less difficult. The score on the first trial was, therefore, discarded when computing the results.

System Maintaining Liquid at Highest Point

In order to maintain the liquid stationary at the highest point for any one trial so that an accurate reading might be obtained, it was necessary to introduce a valve and stopcock. The valve was made of leather and brass and was similar in principle to those used in bicycle pumps. It was inserted in the top outlet of the bulb. When force was applied to the bulb the valve opened allowing the liquid to flow into the tubing. When pressure was removed, the valve closed tightly so that the liquid could not flow back into the bulb at that point.

System Lowering Liquid

About eight centimeters above the valve a bent piece of glass tubing with stopcock was joined to the main tube by a piece of rubber tubing. The other end of this glass tube passed through the baseboard, bent again, and entered the bulb beneath the baseboard. When the stopcock was closed, it together with the valve helped to maintain the liquid at its highest point in the tube. When the stopcock was opened, any liquid from the main tube had to flow through it and drain into the rubber bulb by the route just described, thus lowering the liquid and refilling the bulb.

Lever System

The lever system (Figure 3) consisted of two bars of which the

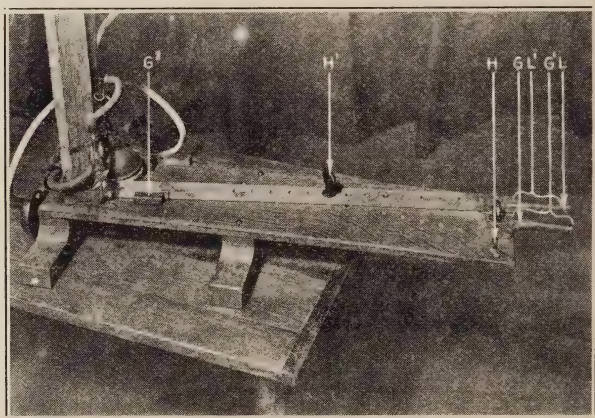


Figure 3 Motivation-Dynamometer Test, Lever System

- H Slot
- G Stationary bar
- L Movable bar
- H1 Screw fastening the movable bar to the board
- G1 Portion of bar grasped by subject
- L1 Grip end of movable bar
- G2 Block of wood acting as stop for bar

one was movable and the other was stationary. The purpose and function of this system was twofold: (1) it was found that squeezing the bulb by direct contact introduced a variable in that not all subjects grasped it at the same point, and that by adding the movable lever bar this variable was eliminated since the bar made contact with the bulb at the same point for every subject and in every trial; (2) squeezing the bulb by direct contact proved to be too easy

even for the two- and three-year-old children. The screw which clamped the movable bar to the baseboard also acted as the fulcrum. When it was moved toward the rubber bulb the task was made less difficult, while moving it to a hole farther away from the bulb made it increasingly more difficult. The stationary bar was $11\frac{1}{2}$ cm. long, 3 cm. wide at widest point, and 1.3 cm. thick. A portion of this bar which was grasped by the subject was hollowed out to fit the hand of the young child. The bar was clamped to the board and was adjustable at any point along a slot. This feature made it possible to adjust the distance between the two bars to the length of the child's palm.

The movable bar was 62.5 cm.; approximately 8 cm. of it made up the grip-portion of the bar. The opposite end of the bar made contact with the rubber bulb. A block of wood acted as a stop for this bar. The grip-end of the bar was slightly hollowed out and was similar to that of the stationary bar. The edges of this portion of both bars were rounded and polished smooth so that there were no sharp edges to cause undue discomfort. The grip-portion of both bars extended beyond the baseboard so that the hand and fingers could move freely in full flexion and extension when they were placed over the two bars. Fourteen holes, each about two centimeters apart, were drilled in this bar.

A screw, which acted as the fulcrum, was made so that the amount of tension for each hole would be the same since the nut could be screwed onto the bolt for a certain distance only.

System for Insuring Success and Failure

One of the necessary requirements of the task was that it permit the experimenter to insure success or failure and at the same time to measure the subject's actual performance. This requirement was met as follows (Figures 1 and 3): A single piece of rubber tubing led from the top of the rubber bulb up the side of the glass tube crossing around to the back at the height of about thirty-six centimeters. Two stopcocks were placed in this piece of tubing. The purpose of the one-way stopcock was to provide a means of allowing air bubbles which rose to this point to be forced out of the system. The two-way stopcock permitted the liquid from the rubber bulb to be shunted into either of two outlets. When the stopcock was turned so as to permit the liquid to go through one outlet it was then shunted up the back tube which was fastened to the upright. This tube was similar to the one on the front of the apparatus and was 4

cm. in inside diameter and 1 m. in length. By this arrangement the experimenter was able to measure the subject's actual performance on the back tube. If the experimenter had wished the subject to see his own performance, it was only necessary to turn the stopcock so that the liquid could go through the other outlet and it was then conducted around to the front tube through the rubber tubing.

In order to insure success or failure and at the same time allow the subject to see the liquid as it went up the tube another system was introduced which the experimenter worked. This is indicated in Figure 3. This system consisted of a rubber bulb, a valve, and a stopcock similar to the one which the child manipulated. The valve and manometer in this system performed the same function as they did in the basic system. This arrangement was introduced into the main system below the stopcock. When the experimenter squeezed her bulb, the liquid was forced into the main system and thus around into the front tube. This bulb was, of course, concealed from the subject. A concrete illustration may help to make the working of this system more clear. The experimenter wished the subject to have a success, that is to see the liquid rise to the top in the front tube so the bell would ring, and yet be able to measure the actual performance of the subject. The stopcock was turned so it was necessary for the liquid from the bulb, which the child manipulated, to go through the outlet and up the back tube. At the same time that the subject started to squeeze, the experimenter began to squeeze the other bulb also. Pressure on this bulb forced the liquid from this system into the main tube and so on around to the front tube through the rubber tubing. The experimenter squeezed the bulb in such a way that the liquid in the front tube advanced at a rate proportional to that in the back tube. If, on the first trial, the subject's reading did not lie somewhere between twenty-five and seventy-five centimeters, the fulcrum was changed making the task more difficult for readings above seventy-five centimeters and less difficult for performances reading less than twenty-five centimeters. Under this arrangement it was necessary for the experimenter to force the water to the top (to insure success) through only a fairly short distance of a few centimeters after the subject had reached his maximum expenditure of energy for that squeeze. In all cases readings on the back tube were taken at the point when the column stopped rising for an appreciable second or so. The characteristic squeeze was one in which the water started rising rather rapidly at first and then more slowly as

the force of energy was finally expended until the column stopped rising altogether. It was, therefore, necessary when the experimenter forced the water in the bulb up the front tube for the short distance remaining, to not do it so rapidly as to be out of all proportion to the first part of the squeeze. For this reason it was especially necessary to adjust the difficulty level of the task so that the readings for the subject would lie between twenty-five and seventy-five centimeters so that the experimenter would not have to force the water up the tube for a long distance after the subject had stopped squeezing. Efficiency in gauging the approximate difficulty level of the subject comes fairly rapidly with practice. A bell which was attached to the back of the apparatus was connected with a dry cell battery. A push button placed in the circuit could be pressed by the experimenter's foot when the water reached the top of the tube.

ADJUSTABLE CHAIR AND ARM REST

The second unit of the apparatus consisted of an adjustable chair and arm rest in which the subject sat while being tested. The purpose of this unit was to stabilize the position of each child in reference to the main system.

Chair

There were two considerations to keep in mind in devising the chair: (1) the length of the chair seat should be narrow enough so that even the youngest subject could sit comfortably with his back against the back of the chair and yet be able to flex his knees over the front of the seat edge without the knees or thighs coming in contact with the cross bars, thus avoiding, if possible, an added mechanical advantage from bracing; (2) the chair must be adjustable, that is raised or lowered in reference to the main system, in order to be able to accommodate the wide age range to be used.

The chair was built as follows (Figure 4): An ordinary automobile jack was fastened to an oak board 26.5 cm. wide, 3 cm. thick, and 30 cm. long. On top of the jack was fastened a small chair with sawed-off legs and narrow seat which had been remade to fit the needs of the two-year-old child. An adjustable step was added to the front of the chair in order to provide a place for the subject's feet. This adjustment was most necessary for the two-, three-, and four-year-old subjects.

Arm Rest

The arm rest was provided in order to insure a uniform stabiliza-



Figure 4 Motivation-Dynamometer Test, Chair and Arm Rest

- 1 Jack for adjusting height of chair
- 2 Chair
- 3 Adjustable step
- 4 Base of arm rest

tion of the arm and hand for all subjects. It was constructed in the following manner: The base of the rest was a smooth board 28 cm. long, 9 cm. wide, and 1.5 cm. thick having two slots on its two outer edges. A back stop, made from a block of wood hollowed out to fit the curve of the elbow, was adjustable along these slots which ran in the longitudinal plane. On the under surface of the base, also running in its longitudinal plane, was fastened a piece of strap iron which was also slotted to provide adjustment for the length of arm from elbow to wrist. This piece of strap iron was con-

nected by means of thumb screws to the slotted upright of the arm rest, and at the same time provided adjustment in the vertical plane. The upright was in turn fastened to an iron bar by means of thumb screws at the side-front of the main chair unit. Adjustable features in the horizontal plane took care of the differences in length of forearm from elbow to wrist of the age range used, while adjustment in the vertical plane took care of the differences in length of arm from shoulder to elbow. A strip of chamois skin about five inches wide was fastened to the inner side of the elbow stop. After the child's arm had been placed in the arm rest this strip of soft chamois was drawn across the arm and fastened on the outside of the elbow rest as seen in Figure 4. This arrangement was designed to keep the subject from lifting the elbow.

TRAIN SYSTEM

If all subjects were to receive success and failure it was necessary to use some means of preventing carry-over from success to failure or vice versa. There would be greater opportunity for carry-over effects to take place if the same apparatus were used for the two conditions. One means of overcoming this difficulty would be to substitute a different piece of apparatus for the second condition. If, however, the second piece of apparatus could make use of the same units of measurement as the first piece, the problem of comparing results under the two conditions would be simplified. Accordingly, a train system was devised in conjunction with the main system to be used for the failure condition. Figure 5 shows this unit from the front as seen by the subject.

This system consisted, in part, of a fan-shaped piece of beaver board which could be fastened to the upright of the main apparatus by means of a hook in such a way that it stood vertically and in front of the upright. On the front of this board, which was painted a neutral tan, were two small red railroad stations made of cardboard. A train track was painted between them. A small toy train, also made of cardboard, was fastened to a rider which moved between the two stations. The mechanism for moving the train could be controlled by the experimenter and was placed on the back of the beaver board, and therefore, not visible to the subject. It consisted essentially of a piston arrangement acting against a coil spring. One end of the spring was fastened to the iron bar-hook which attached the unit to the main apparatus. The opposite end of the spring was wound around the axle. An integral part of the axle

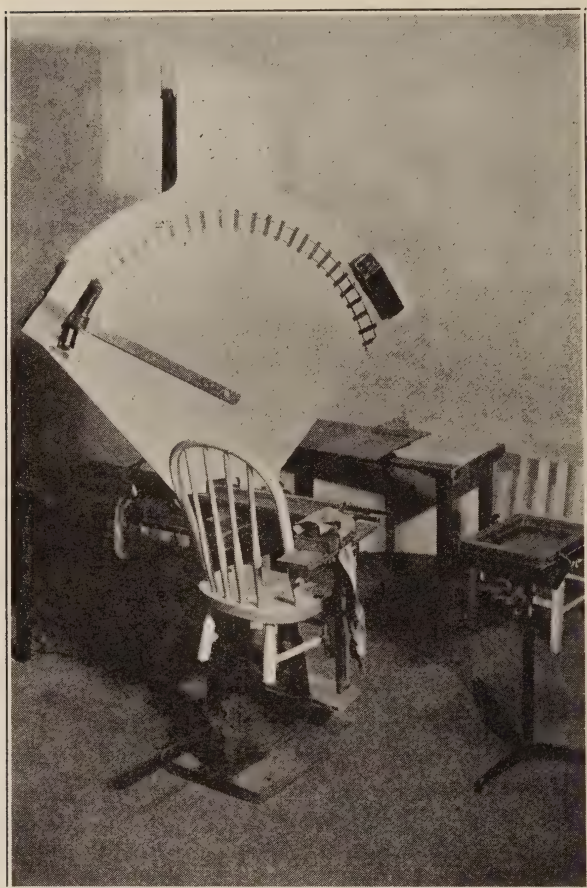


Figure 5 Motivation-Dynamometer Test, Presentation 3 (Train System)

was a pinion, the teeth of which coincided and fitted into similar teeth on the under surface of the piston rod. This piston rod actuated a leather-packed piston head sliding in the brass cylinder. An air vent was inserted in the closed end of the cylinder and consisted of a brass tube into the free end of which was fitted a needle valve and a stopcock. The stopcock was inserted between the air vent and cylinder to allow sudden and abrupt closure of the air passage. In the operation of this mechanism the kinetic energy of the spring rotated the axle and pinion, which in turn forced the piston rod with attached piston into the cylinder compressing the contained air. The speed of the passage of the piston down the

cylinder could be regulated by the extent to which the needle valve was opened and could be stopped abruptly by the closure of the stopcock. The arm which carried the train was mounted on the same shaft as the pinion. Hence, it was possible to regulate the speed at which the train could be moved since the two arms moved in unison.

An illustration will perhaps make the principle clear. Supposing the experimenter wished the subject to see both progress and results of performance, to have a goal to work for, but at the same time also wished, on the one hand, to be certain that the subject would not attain the goal set and yet, on the other hand, wished to be able to measure the actual performance of the subject under these conditions. For this purpose the train system was hooked to the main apparatus and in front of the upright. The child from his position could not see the upright or any part of the main system which he had seen during the second week's testing except the lever bars and board. As in the previous trials the subject was to squeeze the "sticks together" and was told that this time the little train would move on the track. If he (the subject) squeezed hard enough the train would go way over to this station and the bell would ring. At the same time that the subject started to squeeze, the experimenter opened the air vent and by watching the advance of the water up the back column was able to keep the speed of the train proportional to it. The train system was as wide in its horizontal plane as the water system was long in its vertical plane. When the experimenter wished to insure failure it was only necessary to turn off both the stopcock and needle valve to insure full stop of the train arm at any desired point. The distance left to go when the train was stopped to insure failure was kept approximately constant for all subjects and trials in order that the nearness with which the subjects came to attaining the goal would not be a variable. This system was somewhat more difficult for the experimenter to operate than was the water system and considerable practice was needed in order to learn to start and stop easily and to regulate the outlet of air so that the progress of the train was smooth and continuous. In addition, the column of water measuring the child's actual effort had to be watched carefully in order to keep the progress approximately proportional. The position of the lever arm gave the experimenter the exact position of the train arm in front since both moved on the same shaft.

Disadvantages

While the apparatus did allow the experimenter to insure success or failure there were disadvantages, some of which should be corrected before it is used further. From the spontaneous remarks of the subjects it was evident that the manipulation of the "sticks" did cause discomfort. Under success, however, it was noted that those who complained of their hand hurting, in many cases, added, "but I don't care." Under failure this same factor seemed, in some cases at least, to be offered as an excuse for not being able to make the bell ring. A thin strip of felt glued or cemented to the wood might take care of this difficulty somewhat.

The method of calibration which must be used to make all readings comparable is both laborious and tedious and allows for many variables to be introduced when reading and reporting values. It should be checked at least once a week if the apparatus is used continuously. It is probable that some arrangement could be devised whereby energy output could be measured directly at the point of application and thus eliminate the necessity of having to translate readings into equivalent values.

Because the experimenter's attention must be directed on the measurement of the energy expended, it is almost impossible for one person alone to conduct the work if, at the same time, qualitative reactions of the subject are desired.

PERFORATION TEST

The Perforation Test, which was presented to the child immediately following the seven trials on the Motivation-Dynamometer Test, was also a test of motor ability and was devised and used at the Iowa Child Welfare Research Station. Although certain changes were made in the method of presenting and giving this test, the general principle was very similar to that described by Baldwin and Stecher.¹² The task set for the child consisted of punching out the imprint of broken circles in paper, using the blunt end of an orange wood stick. The imprint of the circle was made by punching through a master pattern with a leather hammer punch, the circumference of which had been filed away at two points so that only a broken circle resulted. A light weight ordinary bond paper was used. In the usual procedure the subject was given one minute to see how many

¹² Baldwin, Bird T., and Stecher, Lorle I.: *The Psychology of the Preschool Child*. New York: D. Appleton, 1924. Pp. 305. (p. 95-98)

holes he could punch. Each sheet contained fifty broken circles. It was evident from previous studies on this test that some of the older children in the Junior Primary group at the Iowa Child Welfare Research Station were able to punch out nearly all of the fifty broken circles in one minute and in some cases in less than the time allotted. As the age range of the group used in this study included some seven- and eight-year-old children it was thought advisable to increase the number of holes which could be punched out. A new stencil was made which contained a total of ninety-one broken circles arranged in eleven rows, eight to a row, and in addition three sample circles at the top of the sheet. These samples were used by the experimenter to demonstrate the method of procedure to the subject. Each sheet of paper was held taut between two Montessori frames which were clamped together at each of the four corners. Since the child was required to stand up to do this task, a special table was devised to permit quick adjustment of the working surface to the height of the individual child in question. It was constructed as follows: An oak board about $8\frac{1}{2}$ inches by 11 inches by $1\frac{1}{2}$ inches was fastened to the top end of the sliding member of an adjustable standard which had a steel tripod as its base. A set screw made vertical adjustment a simple matter. The lower Montessori frame used to hold the perforation sheets was permanently fastened to the top of the oak board. Two equalizing clamps, operated by hand wheels, were placed at the ends of the table. Hinged clamps which swung out of the way permitted insertion or removal of paper and top frame when set screws were loosened or tightened by means of the hand wheels. The table could be lowered to about seventeen inches and raised to about thirty-two inches. Adequate for the youngest as well as the oldest subject used, the table proved to be most satisfactory in every way.

MATERIALS OF THE REWARD AND PUNISHMENT SERIES

Under the reward series the subjects were given a star each time they made a success (that is, each time the bell rang). Denison's large gold stars, size No. 4, were used as a reward. A red piece of cardboard 1 inch wide by 7 inches long was marked off into seven equal places, one space for each trial. The subject's name was printed on the front of this card. The subject was told that this card belonged to him and that each time he made the bell ring, the experimenter would put a gold star on his card.

In the punishment series the experimenter used cut-outs of a gingerbread boy. A heavy grade of black drawing paper was used for the forms. Eyes, nose, and mouth were drawn in with white pencil. On one side of the front of what would be the boy's jacket were placed vertically, in a row, seven small red gummed dots designated as red buttons. The subject was told that the little boy belonged to him. His name was also written on the back of the card. The subject's attention was called to the fact that the little boy had buttons down only one side of his jacket, but if he made the bell ring, the experimenter would "put another little red button down this side of the boy's jacket" (indicating side with no buttons); if the bell didn't ring, of course, the experimenter couldn't put a button on the boy's jacket. Instead it would be necessary to cut one of the little buttons off of the boy's jacket. After each trial the experimenter cut off a red dot, using a pair of small scissors. The principles of punishment involved were: of withholding something which the child presumably wanted (that is to get more red buttons) and taking away of something he already had (buttons already on the boy's jacket).

DISADVANTAGES

Both of the recorders as well as the experimenter subjectively felt that the stars served as a true reward to the children but that the little boy was not highly successful as a means of punishment. It may be that punishment is more effective when it becomes more of a personal matter. The subjects perhaps did not truly feel that the little boy was theirs and so they were not upset or disappointed when it was cut up. It would seem that a different picture might have been obtained if, under failure, the stars which the child had won as a reward had been cut off as a "punishment." To have used the stars gained in success series as a reward would have added one more possibility of carry-over from success to failure. The little boy was thus substituted. Because the general principle of withholding as well as taking away some desired thing was not highly successful in this study does not mean that it might not be so in a different set-up. It is perfectly obvious that punishment cannot affect a child unless he realizes that he is being punished. It seemed evident that many subjects did not appear, overtly at least, to feel that they were being punished because many of them smiled or even laughed when the buttons were cut off. It may be that the smiling and laughing was a means of covering their embarrassment. That this was not always the case is indicated by the fact that many, although not all

of them, seemed to be genuinely amused when the little boy was cut up.

RECORDS

A series of nine record blanks was used each time the subject was tested, making a total of twenty-seven blanks for any one subject given the entire number of presentations. Three blanks were used during the first task. The necessary routine information heading each blank was filled out beforehand whenever possible. The first blank, which is not included here, was used by the experimenter to record the subject's energy output in the seven trials given on the Motivation-Dynamometer Test. The second and third blanks (Figures 6 and 7) were filled out by the recorder who checked the items on each blank during the time the child was being tested.

INSTRUCTIONS TO RECORDER

The instructions given to the recorder with these blanks were as follows:

1. Try not to allow your attitude toward any previous child to influence you when recording the response of the child in question.
2. The term response refers not only to the verbal response (which should in all cases be recorded in full) but also to the expressive, overt behavior of the subject. For example, nods head, smiles, wiggles in chair, shrugs shoulders, postural attitude indicative of indifference as well as that of elation or depression, should be noted.
3. When the experimenter enters with the subject she will give the recorder the subject's name. The recorder will find the appropriate folder, take out the second and third blanks, put them on her clip board and place the other six blanks in order on the experimenter's board. This is to be done while the experimenter is adjusting the subject to the apparatus.

IOWA CHILD WELFARE RESEARCH STATION

State University of Iowa

Iowa City, Iowa

Name	Sex	Date of test
		Year Month Day
Group	Condition	Date of birth
		Year Month Day
City	Group	Age
		Years Months Days
State	Series	Recorder

Response of Subject when Approached for Game

- A. What was subject doing when approached?
 B. What was subject's response to "come play game"?

Very Willing | | | Normal | | | Very Unwilling

Response of Subject during Pre-Experimental Period

- A. Response to room To apparatus
 B. Mood as evidenced from entrance until given first directions:

Elated | | | Normal | | | Depressed

Response of Subject during Post-Experimental Period

- A. Response of subject to:
 "That was a nice game wasn't it?"
 "Did you like it?"
 "Would you like to play it again some time?"
 B. Mood as evidenced from end of last trial until beginning of next game:

Depressed | | | Normal | | | Elated

Response of Subject to Second Game

- A. Response of subject to:
 "We are going to play this game now."
 Reluctant | | | Normal | | | Anxious
 B. Response of subject during game:
 Expressive behavior
 Verbal comments
 C. Response of subject at end of game:
 "Would you like to come and play again some time?"

Anxious | | | Normal | | | Reluctant

Figure 6. Record of Behavior Responses to the Experimental Situation

IOWA CHILD WELFARE RESEARCH STATION

State University of Iowa

Iowa City, Iowa

Name Sex Date of test
 Year Month Day
 Group Condition Date of birth
 Year Month Day
 City Group Age
 Years Months Days
 State Series Recorder

Behavior Response Items	Trial							Comments
	1	2	3	4	5	6	7	
A. Expressive Behavior								
1. Contorts face and neck								
2. Breathes hard, face reddens								
3. Lifts elbow, arm								
4. Uses shoulder, back muscles								
5. Attempts to use two hands								
6. Takes extra squeeze								
7. Smiles happily, elated								
8. Smiles matter-of-factly								
9. Smiles shyly, embarrassed								
10. Laughs, with embarrassment								
11. Laughs, matter-of-factly								
12. Laughs happily, elated								
13. Tears, ready to cry								
14. Rejects apparatus								
15. Indifferent								
16. No change								
B. Verbal Comment								
1. Talks irrelevantly								
2. "Likes game"								
3. "Dislikes game"								
4. Rationalizes								
5. Compensates								
6. Compares self with others								
7. Blames apparatus								
8. Blames experimenter								
9. Convinces E. that he played well								
10. Appeals for recognition								
11. Calls attention to work								
12. Asks to leave								
13. Refuses to play								
14. Indifferent								
C. Attention								
1. During trial								
a. On game								
b. Off game								
2. After trial								
a. On game								
b. Off game								
(1) On self								
(2) On environment								
D. Mood During Game								
Elated Normal Depressed								

Figure 7

Behavior
Responses
to
Different
Motivating
Conditions

Behavior Responses to the Experimental Situation

This blank (Figure 6) deals with the response of the subject when approached for the game, during the pre- and post-experimental period, and to the second game. Those responses requiring further explanation than that given on the blank are discussed in this section.

Response of Subject During Pre-Experimental Period.—This includes the time from entrance into the room, during adjustment to chair and apparatus, until first directions are given.

The following instructions are included for response to room and apparatus: Does subject upon entrance to the room make any comment about it? Does he indicate verbally or by expressive behavior any interest in the apparatus? Similarly, is there lack of interest? Indicate kind and degree, if possible.

Under responses indicating mood, the following instructions are given: Along the line given, let normal represent the mid-point as being a matter-of-fact attitude in which the subject neither overtly shows that he is happy (or any degree of that attitude up to elation) or unhappy (or any degree of that attitude down to depression) at having to play the first game. The standard to be kept in mind is that of an adult expression of the attitude.

Response of Subject to Second Game.—This is to be checked beginning with the experimenter's statement that "We are going to play this game now." The degree of the attitude with which the child leaves the first task and turns to the second as shown by the reluctant-anxious rating is estimated and checked on the line. For example, does he take the doing of the second task matter-of-factly, if so, mark it normal; does he appear reluctant (in any degree) to leave the first and play the second; or does he appear anxious (in any degree) to leave the first task for the second? Under "Would you like to come and play again some time?" the principle is the same only it refers to leaving the testing situation and returning to the subject's own room. In making the rating on mood and on the line reluctant-anxious you may place a cross on any part of the line.

During this second game record expressive behavior and verbal comments which help to evaluate the subject's attitude as he performs his task, as for example, sighs, indication of fatigue, glancing back at the first task, absorption in the second task, indications of restlessness, etc.

Behavior Responses to Different Motivating Conditions

The expressive behavior, verbal comment, attention, and mood of the subject were recorded on the blank (Figure 10). Instructions to the recorder are given below.

Expressive Behavior.—Place a 1 after each item which was present during the trial being checked. Place a 0 after each item which was not exhibited in that given trial. In addition, rate Item 1 (contorts face and neck) and Item 7 (smiles happily, elated) on a basis of three points when 1 indicates that the item in question is just noticeable and 3 indicates that it is decidedly noticeable; 2 is anywhere between 1 and 3. (Note that the recorder who checked in the Ames group rated all items under this heading on a basis of three points.)

Verbal Comment.—Under this heading indicate all remarks made from the time directions for the first trial are given until instructions, "We are going to play this game now." Record in full under the blank space for comments. If possible, classify when recording, as for example, "I like this game," "This is fun," etc. Write "Likes game" followed by the comment. Check with a 1 the appropriate square under the trial in question.

Attention.—Usually the direction in which the eyes are focused will indicate this, but not always. If necessary make use of verbal comments of subject in rating attention as, for example, he says, "I like this game," looking toward the experimenter or recorder. If attention is off game (that is the eyes are not focused on the apparatus), check in the appropriate square. If attention seems to be on the game as indicated by focus of the eyes, check in the appropriate square. If, during or after the trial, attention is both on and off the game momentarily, place a 1 in both squares for the trial in question. In addition if attention is off the game is it primarily on self or on environment?

Mood during Game.—Rate mood during the game as was done previously on the first blank. This should be made immediately after the last trial and before the experimenter says, "We are going to play this game now."

Additional records were used during the three trials of the Perforation Test. After the experimenter said, "We are going to play this game now," she took a sheet of paper which contained the broken circles which were to be punched through by the subject and

put it between the frames of the perforation table. After directions had been given, the experimenter followed the pattern of punching holes used by the subject and recorded on a record sheet on which each hole on the child's paper was represented by a dot, numbering each appropriate dot in numerical order to correspond to the position as well as the number of holes which were punched. At every tenth second the experimenter encircled the hole reached at that point in order that a record of the rate of punching for every ten seconds could be obtained. In addition to this record there were certain observational items to be filled in. However, after the experimenter had recorded a number of subjects on all items, as well as pattern and rate, the former record was discontinued since it was impossible, with the older children at least, to accurately record the pattern and rate, if at the same time observational notes had to be taken down. Therefore, in all cases the pattern and rate were obtained without interfering with the main part of the recording.

TEST PROCEDURE

The following section will describe in detail the specific testing procedure which was used from the time the subject was obtained from his school group until he had completed the instructions for the final test series given to him.

METHOD OF OBTAINING SUBJECT FROM HIS GROUP

The method of obtaining a subject varied with the group in question. The children in the kindergartens at Ames and in Grades I and II of the University Elementary School were told that Miss C—— was going to play a game with them some time during the day, and that when they were through playing the game she would tell them whom she wanted to come next. They were asked to see how quickly and quietly they could find that person and tell him that Miss C—— was ready for him. In this way it was not necessary to disturb either the teacher or the routine of the class.

In the preschool laboratories the usual form of procedure for obtaining a child for a "game" was to ask him to "come play a game" or "I have a 'game' to play with you, would you like to come with me now?" or some other similar remark. Usually the experimenter tried not to approach the subject when he was engaged in an absorbing occupation. After gaining his attention the experimenter said, "I would like to have you come with me now. I have something for you to do."

ADJUSTMENT OF SUBJECT TO APPARATUS DURING THE
PRE-EXPERIMENTAL PERIOD

When the experimenter and subject appeared at the door of the examining room, the experimenter gave the recorder the child's name. The recorder then took out the subject's own blanks (of which there were always nine) from the file, placed the second and third blanks on her own clip board, and put the other seven sheets in order on the experimenter's board. During this routine matter the recorder also took note of the subject's reactions during the pre-experimental period as indicated under A and B of this section on the second blank, which included response to the room, response to the apparatus, and mood as evidenced from entrance until given the first directions (Figure 6).

Method of Obtaining Coöperation

As the child entered the room the experimenter waited for a few brief seconds to allow him to become adjusted to the situation and then said, "We are going to sit in this little chair over here. You can get in it yourself, can't you?" If the subject had difficulty or seemed at all timid, the experimenter helped him to get into the chair. There was at first some slight difficulty in getting the younger subjects to approach the chair. From their remarks it was apparent that they could not see just how a chair, whose legs were clear up in the air rather than on the floor where most legs of chairs should be, could stay there without falling down. In order to clear up this point it was necessary, in some cases, for the experimenter to sit in the chair and to point out that it would not fall down. This was usually sufficient to adequately meet the situation. Once the child was established in the chair another slight difficulty sometimes arose. The chair had a tendency to wiggle slightly, especially with the younger subjects whose weight was not great enough to stabilize it. To meet this situation the experimenter, as soon as the child was seated, put one hand on the back of the chair and one on the front of the seat and holding the chair firmly, rocked it back and forth rhythmically, saying as she did so, "John, do you have a little rocking chair at home? (If he did not, the rocker at school was used by way of analogy.) Well, I want to tell you something about this little chair. See it rocks back and forth just like your little rocking chair does at home (or the duck rocker at school). Isn't that fun?" The use of this analogy seemed to link a familiar situation with an un-

familiar one and helped to dispel any uneasiness which might have arisen.

Specific Technique for Adjusting Subject to Apparatus

After the child had established himself in the chair and was at his ease, it was necessary to see that he sat with his back against the back of the chair and with arms at the side. This position should be as natural a one as possible. The foot-rest was then adjusted to the proper position. The right arm was allowed to hang loosely at the side. The height of arm-rest was then adjusted along the slots in the vertical plane so that the base came just to the elbow. The arm was flexed at the elbow until it was at right angles to the body and the forearm was pronated. The experimenter then said to the subject, "Now I'm going to have you put your arm over here. See, we are going to put it clear back and underneath just like you put your arm in your coat sleeve." As the experimenter talked she took the subject's arm firmly and directed it back and underneath the soft chamois band. The distance between the elbow and wrist was estimated and the baseboard and elbow stop were then moved horizontally along the slots until the wrist could be flexed or extended without being stopped by the baseboard. It was necessary to keep in mind that the shoulder and forearm should not be unduly extended during this process. Having completed this adjustment the experimenter said, "Now we are going to pull this band down like this (indicating the chamois band) but we don't want it to be too tight. You tell me if it is too tight, won't you?" When this had been done the chair was pulled forward to the main piece of apparatus, raised or lowered by means of the jack, until the arm rest was just on a level with the grip-bars. The hand and fingers were then placed over the bars or sticks as they were called when referring them to the child. In this adjustment, the wrist and fingers were slightly extended but not more so than is the hand when holding a pencil or pen as used according to the Palmer method of writing. The hand and thumb were not adducted at the wrist but were in a straight line with the arm.

While this procedure may sound formidable and complicated, it should be pointed out that in actual practice many of these steps were combined and once the essential points were learned it did not take more than a minute to a minute and a half to make a complete adjustment for any one subject. One consideration which speeded up this process was that the children, as has been pointed out before,

were fairly homogeneous as to age range for any large school group in which the experimenter was testing. After the major adjustments for the age in question had been determined, it was necessary to make only minor adjustments for the individual child.

DIRECTIONS AND PROCEDURE IN GIVING THE THREE EXPERIMENTAL TEST PRESENTATIONS

After the subject had been adjusted to the chair and apparatus and the experimenter felt that he was comfortable and at his ease, it was possible to proceed with the Motivation-Dynamometer Test. The directions and procedure are given by presentations.

PRESENTATION 1

This presentation was given the first week to all the children the first time they were tested and may be considered as a preliminary control series. Figure 8 shows the apparatus set up for this presentation as seen by the subject. The essential features of this series were that there should be no visual knowledge of progress or results of performance, there should be no mention of an immediate aim (that of making the bell ring), and neither praise nor reproof, reward nor punishment were given since only the directions were to be repeated between the trials. The experimenter attempted to maintain a neutral attitude throughout.

After having adjusted the subject to the apparatus, the experimenter sat down opposite the child and gave the following standard directions: "This is a game to find out how strong you are. See, I put your hand on these two sticks of wood. Now when I say, 'ready, squeeze,' I want you to squeeze the two sticks together as hard as you can." "Ready, squeeze." After each trial the experimenter said, "I want you to try again. Ready, squeeze." The first trial was more or less of an adjustment trial. If, in the case of the younger children, the subject did not seem to understand what was meant by "squeezing the sticks together," the experimenter went around to the subject's position and said, "Watch carefully while I do it. See, I put my hand on the two sticks of wood and when I say, 'ready, squeeze,' I am going to squeeze these two sticks together as hard as I can." (During this explanation the experimenter carried out the instructions.) The experimenter then said, "Now, I want you to do it. When I say, 'ready, squeeze,' I want you to squeeze the two sticks together as hard as you can." "Ready, squeeze." In no case was it ever necessary to show a subject more

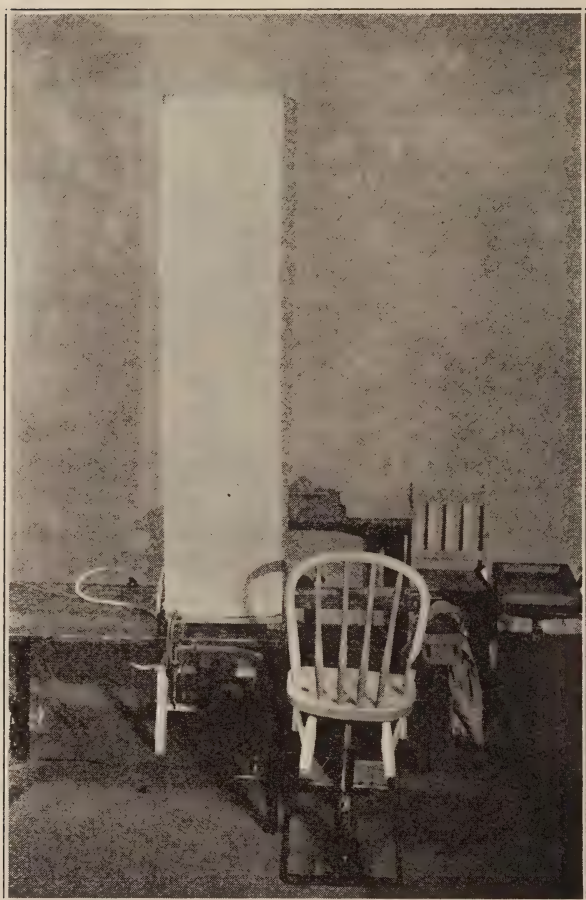


Figure 8 Motivation-Dynamometer Test, Presentation 1

than once just how to squeeze the sticks together. The first trial also acted as an adjustment trial in another way. In obtaining a reading it was desirable that it should lie approximately between twenty-five and seventy-five centimeters if possible. If the task seemed to be too difficult the fulcrum was moved to a hole nearer the grip-portion of the bar, while if it were too easy it was moved nearer the rubber bulb. This was always done after the first trial and the trial then disregarded in working up the data. After the first trial the experimenter unobtrusively moved slightly to the back of the apparatus in order to be able to take the readings from a better position. The experimenter was thus partially screened from view.

Between trials the short directions, "I want you to try again. Ready, squeeze," were given in a matter-of-fact voice.

After the first directions had been given, the recorder transferred her attention to the third blank in order to record the responses of the subject during the seven trials given on this dynamometer. At the end of the seven trials the experimenter asked these questions: "That was a nice game, wasn't it? Did you like it? Would you like to play it again some time?" It is questionable whether the answers to these questions are significant. A child seemed to nod in the affirmative because it seemed to be the thing that was expected of him. Only a very few children said that they did not like the game. It is difficult to get the child to really say what he thinks about a situation. Of course, all spontaneous conversations were recorded and these do in a measure give a check to these questions. It may have very well been the case that the question, "That was a nice game, wasn't it?" may have from the beginning created this attitude. During the later part of the experiment this question was not asked. The experimenter merely said, "Did you like that game? Would you like to play it again some time?" After these responses were given the experimenter said, "We are going to play this game now."

PRESENTATION 2

After the subjects had been given the first test they were assigned to one of four groups according to the procedure described on page 33. Group A thus received Presentation 1 in the second week's testing, given as has been described above. Since this procedure was the same in all three weeks it will not be described again. Groups B, C, and D were given success on Presentation 2 in the second week. Success was defined as the "objectification and realization of a mental aim or end," that is, success means completing some act of behavior which you have set for yourself. In this study it refers "to making the bell ring." When a subject in Group B, C, or D came in he found the apparatus set up as in Figure 9. The screen had been removed and the main apparatus was thus exposed to the view of the subject. After the child had been adjusted to the apparatus the experimenter gave the following standard directions: "This is a game to find out how strong you are. See, I put your hand on these two sticks of wood. When I say, 'ready, squeeze,' I want you to squeeze the two sticks together as hard as you can. Now when you squeeze the sticks together this time, this pretty red

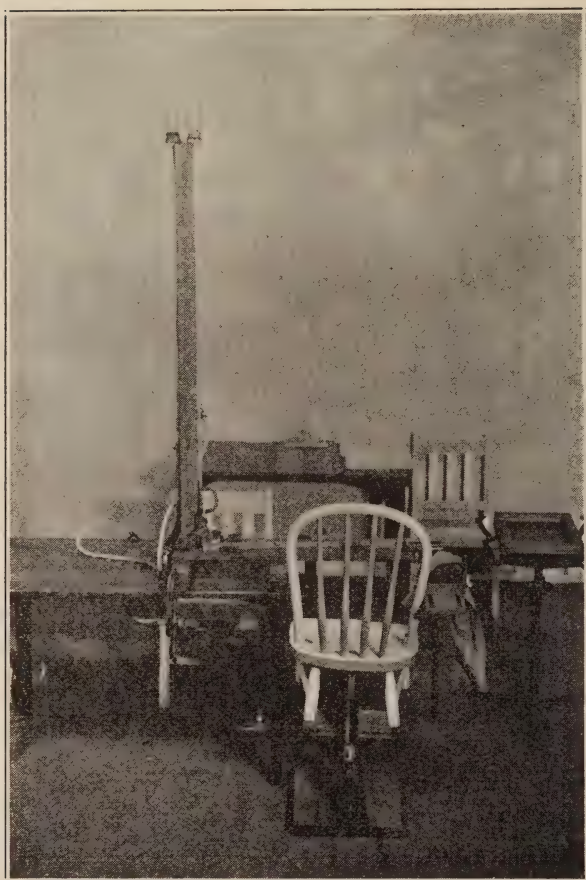


Figure 9 Motivation-Dynamometer Test, Presentation 2

water will go up this tube.” (Indicate each by pointing, being sure that the experimenter has the subject’s attention.) “If you squeeze hard enough the water will go way to the top and a bell will ring. Of course, if you don’t squeeze hard enough, the water won’t go to the top and the bell can’t ring, can it?”

In addition to the condition success, two of the three groups received an added incentive. Thus Group B was designated as the repetition group. After each trial the experimenter said, “I want you to try again.” “When I say, ‘ready, squeeze,’ I want to see if you can make the bell ring.”

Group C was designated as the praise group. The child was praised after each trial. Because so many trials were given it was quite ob-

vious that to repeat the same phrase seven different times would border on the ridiculous. To remove this difficulty a standardized phrase was chosen to give after each trial and each phrase was different. However, all subjects were given the same praise in the same order.

The following standard phrases were given after the appropriate trial:

1. "Splendid, ...! You certainly must be a big, strong to make the bell ring."
2. "Good for you. That was fine that time."
3. "Why, ..., I'm so pleased at the way you're doing."
4. "That's fine. You're doing very well."
5. "You know, ..., I'm certainly proud of the way you're doing."
6. "That's fine. You've made the bell ring every time, haven't you? You've certainly done very well."

The questions on the blank were then asked followed by the Perforation Test.

Group D was designated as the reward group. In this series the same general directions were given and in addition: "See this pretty red card? It belongs to you, see, here is your name on it (indicate name). Now every time you make the bell ring I'm going to put one of these pretty gold stars on your card. Of course, if you don't make the bell ring I can't put a star on your card, can I?" After each trial the experimenter said, "The bell rang that time, so I'll put another star on your card." At the end of the experiment for that day if the subject asked to take his card with him, the experimenter said, "Miss C—— is going to keep your card until you come back to play the game again. Then she will give it to you and you may take it home if you wish." Or, "Miss —— (teacher in group) will give it to you later." This was done in order to keep the children from taking their cards into the school situation and showing them to the other children who might not have had the test yet or who were not in this group.

PRESENTATION 3

One week later each subject was seen again for the third week's testing. Group A was again given Presentation 1. Groups B, C, and D were given failure on Presentation 3 (Figure 5). Conversely to success, failure means "lack of realization or objectification of a mental aim or end," or in this case not making the bell ring. In

addition, each group was given the following added incentive: Group B was given merely repetition of directions between trials. After each trial the experimenter said, "I want you to try again. When I say, 'ready, squeeze,' I want to see if you can make the bell ring."

Group C was this time reproved for not making the bell ring. As in the praise series, a standard phrase was given after each different trial, as follows:

1. "Why, ..., I thought that you were a big, strong, but I guess you're not. You didn't make the bell ring, did you?"

2. "That wasn't good at all. You didn't do at all well that time."

3. "Why, ..., I'm not at all pleased with the way you are doing."

4. "You're not doing at all well. That wasn't good that time."

5. "You know, I'm disappointed in the way you're doing."

6. "You haven't made the bell ring at all, have you? You certainly haven't done at all well."

The questions on the blank were then asked followed by the Perforation Test.

This time Group D was given punishment (withholding something that a child desired and taking away something he already had). In addition to the general directions the experimenter also said, "See this nice little boy? He belongs to you, see, there is your name (indicating name on back of card)." "You see the little boy has pretty red buttons on this side of his jacket, but he hasn't any down this side. Now every time you make the bell ring I am going to put a little red button on your little boy's jacket. Of course, if you don't make the bell ring I can't put a red button on your little boy's jacket, can I? I'll have to cut one of these little buttons off."

PROCEDURE DURING PERFORATION TEST

After each presentation, three trials on the Perforation Test (second task) were given. After Task I was completed the experimenter said: "You may get down and come over here to this little table." The sheets for the Perforation Test were then inserted in the frames and clamped down. The subject was asked to stand on the short side of the table and to the experimenter's right. The experimenter then said, "Do you see all of these marks on the paper?" (The experimenter pointed to the broken circles.) The experimenter continued, "This is called a punch (indicating orange wood stick). When I take the punch and find a mark and punch it like

this (punching as the directions were given) it makes a hole. Now I'm going to give you the punch and when I say, 'ready, begin,' I want to see how many holes you can make before I tell you to stop." The punch was then handed to the subject, and the stop watch was started as the child punched out the first hole. Some subjects would punch out a first hole and then look up for approval. The experimenter said in a matter-of-fact voice, "That's right, now keep right on punching. I want to see how many holes you can make before I tell you to stop."

As the child punched each hole the experimenter followed the subject's pattern of punching on a duplicate sheet where eighty-eight marks were arranged, in the same order as were the circles on the master sheet. In order to facilitate recording, the marks were divided into four quadrants in such a way that one line divided the series between the fourth and fifth line reading across the page and between the fifth and sixth line reading down the page. A certain number of items which were indicated on the blank were filled in when possible. However, in the present study this procedure was discontinued except when possible to do so without interfering with the main recording. Since it required the experimenter's undivided attention to number the correct dots on her sheet in order to correspond with the ones on the child's sheet, the holes punched at the tenth, twentieth, thirtieth, fortieth, fiftieth, and sixtieth seconds were encircled.

After the last trial the experimenter said, "We're all through for today and we're ready to go back. Would you like to come again some time?" If the subject said, "Yes," the experimenter then said, "When you come back again, which game would you like to play?" If the child did not spontaneously signify a choice the experimenter said, "This game" (pointing to one) or "This game" (pointing to the other). The experimenter alternated the game given first with every other child in order that there would be no constant influence of the first or last game being chosen. With the older children it was enough to say, "Which game did you like the better (or 'the best' as this seemed to be understood better than the comparative)." The responses were taken down by the recorder. In case the subject said, "I like both," the experimenter then said, "If you could only play one of the games again, which game would you like to play the most?" Both replies were recorded and the preference underlined. It was probable in some cases that a child,

even though he said he liked one or the other of the "games," did not definitely make a choice but merely said something because it seemed to be expected of him to do so. In most cases, however, the subjects seemed to make a thoughtful reply, and in outstanding cases other remarks previous to these questions indicated that their choices were true choices. On the whole, we may probably consider that these replies were fairly indicative of a subject's true choice.

After the subject had indicated which game he liked best, he was allowed to return to his room. The average time for the total test on any one day was ten minutes.

CHAPTER V

ANALYSIS OF EXPERIMENTAL RESULTS

The results of this investigation have been studied (1) relative to the comparability of the groups on the basis of initial performance on the Motivation-Dynamometer Test; (2) the influence and the effectiveness of the different types of incentives upon the performance of the Motivation-Dynamometer Test; (3) the influence of incentives upon performance in the Perforation Test; (4) the reliability of both tests; and (5) certain qualitative aspects of the effects of motivation.

COMPARABILITY OF GROUPS

In order to ascertain the relative effectiveness of the different types of external incentives which were being studied, it was necessary to have at least four groups if practice, fatigue, and carry-over effects which might be operating were to be taken into account. It has been pointed out in a previous chapter that it was not possible to equate these four groups on the basis of performance in the initial test before any other variable was applied. For this reason it was necessary to assign a subject to a test group on a purely random basis. This was done by listing the subjects in their school groups in alphabetical order and numbering them accordingly. A pack of cards was similarly shuffled, and dealt into four piles constituting Group A, Group B, Group C, and Group D. A subject was, therefore, automatically put into the group in which his number fell.

The following tables indicate the results as to the comparability of the groups according to age, sex, means, and standard deviations of the average of the last six scores on Series I of the Motivation-Dynamometer Test for all subjects who completed all of the three test presentations.

The distribution¹³ of children by groups on Series I of the Motivation-Dynamometer Test according to sex, number of children, mean age, and age range is given in Table 1.

¹³ In computing the results on the Motivation-Dynamometer Test the first trial was disregarded since it was an adjustment trial and, therefore, not comparable for all subjects.

Two hundred thirteen children, including 114 boys and 99 girls, completed all three test presentations. There were in each group

Table 1
Number of Children, Mean and Range in Age in Months
by Groups in Series I for Boys and Girls

Sex Groups	Children	Mean, Age	Range, Months
Group A			
Boys	25	67.88	38-86
Girls	25	61.76	28-92
Boys and girls	50	64.82	28-92
Group B			
Boys	32	65.34	32-94
Girls	25	64.68	34-96
Boys and girls	57	65.01	32-96
Group C			
Boys	27	69.67	34-92*
Girls	26	58.31	29-78
Boys and girls	53	63.99	29-93*
Group D			
Boys	28	64.82	31-90
Girls	23	67.00	46-95
Boys and girls	51	65.91	31-95
Combined Groups			
Boys	112	66.93	31-94*
Girls	99	62.94	28-95
Boys and girls	211	64.93	28-96*

* Two extreme cases in Group C were omitted in making the groups more nearly comparable in terms of mean and standard deviation. See page 74 for explanation.

more boys than girls, and in all but Group D the boys were on the average older than the girls. As a whole, however, the groups were fairly similar as to total number of subjects, mean age, and age range. There was one exception in the case of the average age of the boys in Group C, about which more will be said later.

The mean time (in days) elapsing between the giving of Series I and Series II, Series I and Series III, and Series II and Series III for subjects according to groups was:

Group	Series I and II	Series I and III	Series II and III
Group A	7.06	14.04	6.98
Group B	6.85	13.98	7.13
Group C	7.28	14.34	7.06
Group D	7.59	14.74	7.15

Table 2

Number of Children, Means, in Kilograms, and Standard Deviations
of Scores of Subjects, by Groups on Series I, II, and III
on the Motivation-Dynamometer Test

Group	Children	Mean, Kilograms	Probable Error of Mean	Standard Deviation	Probable Error of Standard Deviation
Series I					
Group A ₁	50	6.430	.292	3.060	.207
Group B ₁	57	6.395	.248	2.770	.175
Group C ₁	53	6.896	.278	3.005	.197
Group D ₁	51	6.598	.261	2.760	.184
Series II					
Group A ₂	50	6.280	.279	2.928	.198
Group B ₂	57	8.184	.308	3.445	.218
Group C ₂	53	8.877	.341	3.783	.248
Group D ₂	51	9.794	.364	3.847	.257
Series III					
Group A ₃	50	6.220	.227	2.899	.196
Group B ₃	57	8.991	.309	3.460	.219
Group C ₃	53	10.462	.398	4.296	.281
Group D ₃	51	10.225	.413	4.424	.295

¹⁴ Throughout this chapter the following notation is used:

Series I (first week testing)

Group A₁ = Control-motivation

Group B₁ = Control-motivation

Group C₁ = Control-motivation

Group D₁ = Control-motivation

Series II (second week testing)

Group A₂ = Control 1

Group B₂ = Success repetition

Group C₂ = Success-praise

Group D₂ = Success-reward

Series III (third week testing)

Group A₃ = Control 2

Group B₃ = Failure-repetition

Group C₃ = Failure-reproof

Group D₃ = Failure-punishment

The means and standard deviations of the average scores by groups for subjects on Series I,¹⁴ II, and III on the Motivation-Dynamometer Test are given in Table 2.

The actual differences in means between Group A, Group B, Group C, and Group D on Series I range from .085 to .778 and from .01 to .516 in the case of the standard deviations. The ratios of the actual differences to the probable errors of the differences of the means of Group A with B, Group A with D, and Group B with D ranged from .22 to .57. In the case of the largest ratio there were approximately sixty-six chances in one hundred that the true differences were greater than zero. However, when the ratio differences in the case of Group C were compared with any other group they were found to vary from 1.34 to 2.01. Since a ratio of 2.01, although not statistically significant, indicates that there are ninety-one chances in one hundred that the true differences are greater than zero, and since all other ratios between this group and the other groups were one probable error or more, there is some possibility that Group C might not have been drawn from the same population as Group A, Group B, or Group D.

The ratio differences of the standard deviations follow somewhat the same trend. In two cases where the standard deviation of Group C was compared with those of the other three groups, they were found to be as high as 1.85, indicating that there are eighty-nine chances in one hundred that the true differences are greater than zero.

It was found that the mean and standard deviation of Group C were both higher than those of any other group. An inspection of the frequency distribution of the scores for the subjects in this group showed that three extreme cases at the upper limits of the distribution were evidently affecting both the mean and the standard deviation. Moreover, in no other group was there more than one such extreme case. Each of the three cases were boys and since, as has been pointed out above, the mean age of the boys of this group was higher than that for any other group, the two oldest subjects of the three extreme cases were omitted from the computations. The mean age of the boys of Group C was thus lowered from 71.20 to 69.67 months. The mean of the average scores of Group C was lowered from 7.173 to 6.896, and the standard deviations from 3.276 to 3.005. The ratio differences of the means of all groups now vary from .22 to 1.34 indicating that, in the case of the largest ratio, there are but

eighty-two chances in one hundred that the true differences are greater than zero. Similar ratios for the standard deviations between all groups now range from .04 to 1.09, showing that there are but seventy-seven chances in one hundred, in the case of the largest difference, that the true differences are greater than zero.

The four groups could now be considered comparable as judged by their performance on Series I of the Motivation-Dynamometer Test. In all further computations, therefore, the two extreme cases in Group C were omitted.

EFFECTIVENESS OF INCENTIVES ON MOTIVATION-DYNAMOMETER TEST

SERIES II

On Series II of the Motivation-Dynamometer Test, Group A acted as a control group and was given the test under the same conditions as under Series I. Groups B, C, and D were given success on Presentation 2. Between trials Group B was merely asked to repeat the test, Group C was praised, and Group D was rewarded with a star for "making the bell ring."

On Series II the means and standard deviations of Groups B, C, and D are all higher than those of Group A. The formula¹⁵ used for computing the reliability of these differences is:

$$PE_{(diff.)} = \sqrt{PE^2_{(av.1)} - PE^2_{(av.2)}}$$

The reliability of the differences of the means and standard deviations of the scores by groups on Series I, II, and III, and Series I and II, Series I and III, and Series II and III on the Motivation-Dynamometer Test is given in Table 3.

The ratios of the actual differences to the probable errors of the differences of the mean of Group A₂ with the means of Group B₂, C₂, or D₂ ranged from 4.58 to 7.67. Since a ratio difference of four probable errors indicates absolute certainty that the true difference is greater than zero, it is evident that these differences are statistically significant. Between success-repetition and success-reward (Group B₂ with Group D₂) there was a ratio difference of 3.38, indicating that there are ninety-nine chances in one hundred that the true differences are greater than zero. While this difference does not indicate absolute certainty it is very probably indicative of a true difference. The ratio difference between the success-repetition

¹⁵ Garrett, Henry E.: *Statistics in Psychology and Education*. New York: Longmans, Green, 1926. Pp. xiii, 317. (p. 133)

and the success-praise group (Group B₂ with C₂) was 1.48, showing that there are only approximately eighty-four chances in one hundred that the true difference is greater than zero. The reliability of the difference between the success-praise group and the success-reward group (Group C₂ with D₂) was $.917 \pm .505$. There are only approximately eighty-nine chances in one hundred that the true difference is greater than zero. Neither of these last two differences are, therefore, statistically reliable. However, the chances are ninety in one hundred that the true mean of the success-praise group (8.877) is greater than the obtained mean of the success-repetition group (8.18); while the chances are ninety-four in one hundred that the true mean of the success-repetition group would be less than the obtained mean of the success-praise group.

When we consider the means of the success-praise group (8.88) and success-reward group (9.79), we find that there are ninety-six chances in one hundred that the true mean of the praise group is less than the obtained mean of the reward group; while there are ninety-five chances in one hundred that the true mean of the reward group is greater than the obtained mean of the praise group.

SERIES III

On Series III of the Motivation-Dynamometer Test, Group A again acted as a control group and was given the test under the same conditions as under Series I and II. Groups B, C, and D were given failure on Presentation 3 of the test. Between trials Group B was asked to repeat the test, Group C was reproved, and Group D was punished for "not making the bell ring."

On Series III the means and standard deviations of Groups B, C, and D are all higher than those of Group A.

The ratio differences between the means of the control and the failure-repetition groups, the failure-reproof, and the failure-punishment groups were all statistically significant since the ratios of the differences to the probable errors of the differences are six probable errors or more, thus insuring with absolute certainty that the true differences are greater than zero.

The reliability of the difference between the means of the failure-repetition and the failure-reproof groups was $1.47 \pm .503$. There are approximately ninety-seven chances in one hundred that the true difference is greater than zero. The reliability of the difference between the means of the failure-repetition and the failure-punishment groups gave a ratio of 2.37, indicating that there are approx-

imately ninety-four chances in one hundred that the true differences are greater than zero. While neither of these differences indicate absolute certainty, it is highly probable that the first is a true difference while the second approaches significance. The chances are ninety-nine and seven-tenths in one hundred, or almost absolute certainty, that the true mean of the failure-repetition group (8.99) is less than the obtained mean of the failure-reproof group (10.46) or the failure-punishment group (10.225). There are ninety-nine chances in one hundred that the true mean of the failure-reproof group is greater than the obtained mean of the failure-repetition group.

The ratio difference between the failure-reproof and the failure-punishment groups was .41. The chances that the true difference is greater than zero are about sixty-one in one hundred. This difference is not statistically reliable.

SERIES I AND II

The reliability of the differences of means and standard deviations of work done (in kilograms) in Series I with the amount of work done in Series II on the Motivation-Dynamometer Test are given in Table 3.

Table 3

Reliability of Differences of Means and Standard Deviations of Scores by Groups in Series I, II, and III, and Series I and II, Series I and III, and Series II and III on the Motivation-Dynamometer Test

Groups	Difference in Means	Probable Error of Difference	Ratio of Difference to Probable Error of Difference	Difference in Standard Deviations	Probable Error of Difference	Ratio of Difference to Probable Error of Difference
Series I						
Group A and Group B	.085	.382	.22	.293	.272	1.08
Group A and Group C	.693	.417	1.66	.213	.295	.72
Group A and Group C	.416	.404	1.03	.058	.286	.20
Group A and Group D	.118	.391	.30	.303	.277	1.09
Group B and Group C	.778	.387	2.01	.506	.274	1.35
Group B and Group C	.501	.373	1.34	.235	.265	.89
Group B and Group D	.203	.359	.57	.010	.255	.04
Group C and Group D	.575	.396	1.45	.516	.279	1.85
Group D and Group C	.298	.382	.78	.245	.270	.91

Table 3 (continued)

Reliability of Differences of Means and Standard Deviations of Scores by Groups
in Series I, II, and III, and Series I and II, Series I and III, and
Series II and III, on the Motivation-Dynamometer Test

Groups	Difference in Means	Probable Error of Difference	Ratio of Difference to Probable Error of Difference	Difference in Standard Deviations	Probable Error of Difference	Ratio of Difference to Probable Error of Difference
Series II						
Group A ₂ and Group B ₂	1.904	.416	4.58	.517	.295	1.75
Group A ₂ and Group C ₂	2.597	.448	5.80	.855	.318	2.69
Group A ₂ and Group D ₂	3.514	.458	7.67	.919	.324	2.84
Group B ₂ and Group C ₂	.693	.467	1.48	.338	.322	1.02
Group B ₂ and Group D ₂	.610	.476	3.38	.402	.338	1.19
Group C ₂ and Group D ₂	.917	.505	1.82	.064	.358	.18
Series III						
Group A ₃ and Group B ₃	2.771	.415	6.68	.561	.293	1.91
Group A ₃ and Group C ₃	4.242	.485	8.75	1.397	.342	4.08
Group A ₃ and Group D ₃	4.005	.502	7.98	1.525	.354	4.31
Group B ₃ and Group C ₃	1.471	.503	2.92	.836	.356	2.35
Group B ₃ and Group D ₃	1.234	.520	2.37	.964	.367	2.63
Group C ₃ and Group D ₃	.237	.577	.41	.128	.407	.31
Series I and II						
Group A ₁ and Group A ₂	.200	.114	1.75	.135	.077	1.75
Group B ₁ and Group B ₂	1.789	.176	10.16	.675	.130	5.19
Group C ₁ and Group C ₂	1.981	.184	10.77	.778	.130	5.98
Group D ₁ and Group D ₂	3.196	.126	25.37	1.087	.095	11.44
Series I and III						
Group A ₁ and Group A ₃	.260	.126	2.06	.164	.084	1.95
Group B ₁ and Group B ₃	2.596	.195	13.31	.690	.145	4.76
Group C ₁ and Group C ₃	3.566	.249	14.32	1.291	.179	7.21
Group D ₁ and Group D ₃	3.627	.257	14.11	1.664	.181	9.19
Series II and III						
Group A ₂ and Group A ₃	.060	.100	.60	.029	.063	.46
Group B ₂ and Group B ₃	.807	.155	5.21	.015	.110	.14
Group C ₂ and Group C ₃	1.585	.179	8.85	1.291	.179	7.21
Group D ₂ and Group D ₃	.431	.167	2.85	.577	.114	5.06

In all cases the reliability of the differences of the means on Series II over those on Series I was statistically significant except

for Group A. The slight decrease in mean for this group is not reliable since there are eighty-eight chances in one hundred that the true difference is greater than zero.

SERIES I AND III

The reliability of the difference of the means and standard deviations of scores made in Series I with those made in Series III on the Motivation-Dynamometer Test are given in Table 3.

The slight decrease in the mean of Group A between Series I and Series III is not statistically reliable ($.260 \pm .126$) since there are ninety-one chances in one hundred that the true mean is greater than zero. The ratio differences between the means of Group B, Group C, and Group D on Series III over those on Series I are all statistically significant.

SERIES II AND III

The reliability of the difference between Group A on Series II and Series III was $.060 \pm .100$ (Table 3). This difference is not statistically reliable since there are only sixty-six chances in one hundred that the true difference between Group B on success-repetition and failure-repetition was $.807 \pm .155$. This difference is statistically reliable. The reliability of the difference between Group C on success-praise with failure-reproof was $1.59 \pm .179$. This difference is significant since the ratio difference was 8.85. The ratio difference of 2.85 between Group D as given success-reward and failure-punishment approaches significance since there are ninety-seven chances in one hundred that the true difference is greater than zero.

INTERPRETATION OF RESULTS

Series II

The results on the effectiveness of the types of incentives given in Series II on the Motivation-Dynamometer Test indicated that success-repetition, success-praise, and success-reward were all more effective for performance on this task than control-motivation. If the subject had knowledge of progress and results of performance, an aim to work for (to make the bell ring), and attainment of the goal, more energy was expended. It is not possible from the results of this study to ascertain which is more effective for performance of a task, visual knowledge of progress and results or having an aim to work for. In order to have answered this question it would have

been necessary to have given to another comparable group of subjects visual knowledge of progress and results, but no aim. It is certain, however, that a combination of both of these two types made for more effective performance on this task.

It is highly probable that more energy is expended if a subject is rewarded with a star for attaining the goal set (success-reward) than if he is merely asked between trials to repeat the task (success-repetition).

While the reliability of the differences between the means of the success-repetition and success-praise groups was not statistically significant, there was evidence to indicate that it would at least be a safer procedure to praise a child for attaining the goal than not to praise him. Likewise, although the differences between success-praise and success-reward were not statistically significant, the chances of obtaining a better performance by rewarding a child are greater than if praise, as given in this investigation, were used.

It should be noted that in obtaining the reliability of the differences between groups it is necessary to use the probable error of the difference formula¹⁶ for uncorrelated measures because the variables are given to different groups. Thus it is not possible to ascertain whether performance under success-praise, for instance, is correlated with that under success-reward. As will be pointed out in a later section, we do know that there is fairly high correlation between performance of the same subjects under success-repetition with performance under failure-repetition, etc. Therefore, it is probable that performance under success-praise and success-reward if given to the same group might be correlated. If this is true the differences which do exist may very well be significant.

In evaluating the influence of praise upon performance it is possible that its effectiveness lies not so much in what is said as in how it is given. In an effort to obtain some objective check on how the experimenter was giving this type of motivation three photographic records of the phrase "Splendid, John, you certainly must be a big strong boy to make the bell ring" were taken on Metfessel's (35) strobophotograph.¹⁷ The top record in Figure 10 was taken first. Thirteen days later the other records in Figure 10 were taken within

¹⁶ Holzinger, Karl J.: *Statistical Methods for Students in Education*. New York: Ginn, 1928. Pp. 372. (p. 243)

¹⁷ Appreciation is due Dr. Harold M. Williams, Iowa Child Welfare Research Station, and Dr. J. H. Tiffin, Department of Psychology, State University of Iowa, for their help in obtaining and interpreting these records.

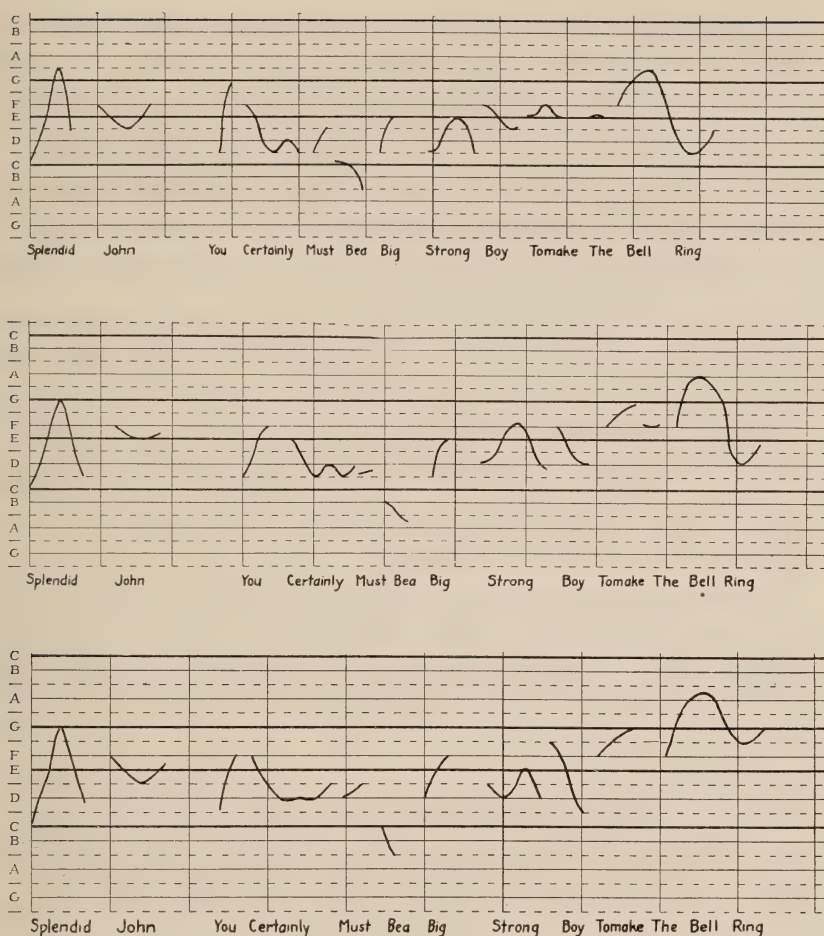


Figure 10 Motivation-Dynamometer Test, Strobophotographic Records

one-half hour's time. They show the pitch and time characteristics of the sentence as read. Comparison of the three records shows a fairly high consistency in the amount of total time, the rhythmic spacings of elements, and the pitch inflections. This type of record, however, gives no direct indication of quality and intensity except in so far as they may be correlated with inflection. Emphasis, quality, and intensity are probably as important if not more so than pitch or time elements in the effectiveness of either praise or reproof inflectives covering a large pitch range, in contrast to small ones in themselves conferring emphasis. Since accent or emphasis is a complex in which all four qualities take part it is probable that pitch

and intensity parallel each other to a considerable extent. When the phrases were learned by the experimenter it was found useful to underline or otherwise mark the words which were to be emphasized.

The experimenter's voice is not that of a trained elocutionist and in many respects it is probable that it was not used as effectively for praise or reproof as might have been the case. An effort was made, however, to give the phrases in an ordinary tone of voice and with a certain amount of inflection and emphasis which would naturally be given to the child in an everyday situation.

However, even though the investigator gave the phrases consistently and with about the same general quality, intensity, pitch, and time characteristics, it should be pointed out that different results as to the effectiveness of this type of motivation might be obtained by another person giving these same phrases, due to the intrinsic difference in emphasis, accent, or quality. Certainly further investigations should be made with young children to ascertain how they respond to such intrinsic differences.

If, however, another person should wish to repeat this study using this type of motivation or make comparisons with it, it would be possible to do so by using the graphs as a basis for learning the pitch and time characteristics of the phrase as given by the investigator of this study.

Series III

From the results on the effectiveness of the types of incentives given in Series III it is evident that failure-repetition, failure-reproof, and failure-punishment were all more effective for the performance of the Motivation-Dynamometer Test than was control-motivation. More energy was expended by the subject if he had visual knowledge of both progress and results, was given a goal to work for even though he did not succeed in attaining that goal and, in addition, was either reproofed or punished for his failure to attain it.

It is also probable that either reproving or punishing a subject is more effective than merely asking him to repeat the test when he has not succeeded in attaining the goal set. The chances of obtaining a better performance under either reproof or punishment as given in this study were greater than if they were not used. Since the difference between failure-reproof and failure-punishment was not statistically reliable, it may be concluded that reproof and punish-

ment as used in this study are about equally effective for performance of the task if the subject has failed to attain the goal.

It is necessary to keep in mind that the formula for obtaining the probable errors of the differences is the one for uncorrelated measures (See footnote page 80) since it is impossible to ascertain whether or not there is a relationship between what a subject does under failure-repetition with what he does under failure-reproof and failure-punishment.

Series I and Series II

These results show that success-repetition, success-praise, and success-reward are all more effective for performance in this test than control-motivation. It is highly important to give the subject visual knowledge of progress and performance and an aim toward which to work.

Series I and Series III

These results indicate that when a goal has been set, and when knowledge of progress of results is given, performance will be more effective than if the subject has no aim and is given no knowledge of progress of results even though he fails to attain that goal and is in addition reproved or punished for not attaining it.

Series II and Series III

In general, then it would seem that failure-repetition, failure-reproof, and failure-punishment are more effective for performance than are success-repetition, success-praise, and success-reward. This statement must be qualified in three respects: the increased performance may be due in part to either having had a previous success at another time with the same experimenter in a somewhat similar situation, or to greater interest in the train system than in the water system, or to a combination of both of these factors.

In order to ascertain the carry-over effects from success to failure it would be necessary to give a group failure on the train system first, and success on the water system last. To ascertain the influence of the interest value of the two pieces of apparatus it would be necessary to give failure on the water system and success on the train system.

Until this procedure has been carried out it is only possible to say that if a subject has had a previous success with the experimenter, and then comes back to a similar situation which may be more in-

teresting to him, he will exert a greater amount of energy even though he is failing to attain the goal set than he did when he was succeeding.

In evaluating these results it is necessary to keep in mind that it is possible that at least two other factors than failure itself may be influencing the results: (1) greater interest in the train system, and (2) a previous success with the same experimenter in a somewhat similar situation. More will be said about these two factors when the relative effectiveness of success and failure are compared directly.

In computing the results on the effectiveness of the incentives as given in Series I with those in Series II, Series I with Series III, or Series II with Series III, the formula for the probable error¹⁸ of the differences for correlated measures was used:

$$P.E.M_1-M_2 = \sqrt{(P.E.M_1)^2 + (P.E.M_2)^2 - 2r_{12} P.E.M_1 P.E.M_2}$$

The intercorrelations in terms of Pearson product moment correlation for each of the three series and for each of the four groups are:

Groups	Children	r	P.E.
Group A ₁ with Group A ₂	50	.929 ± .013	
Group A ₁ with Group A ₃	50	.900 ± .018	
Group A ₂ with Group A ₃	50	.939 ± .011	
Group B ₁ with Group B ₂	57	.82 ± .029	
Group B ₁ with Group B ₃	57	.767 ± .040	
Group B ₂ with Group B ₃	57	.87 ± .022	
Group C ₁ with Group C ₂	53	.85 ± .026	
Group C ₁ with Group C ₃	53	.786 ± .036	
Group C ₂ with Group C ₃	53	.890 ± .017	
Group D ₁ with Group D ₂	51	.967 ± .006	
Group D ₁ with Group D ₃	51	.813 ± .032	
Group D ₂ with Group D ₃	51	.92 ± .015	

VARIABILITY

In order to ascertain the effects of motivation upon variability the coefficients of variation and their probable errors were computed. These together with the means, the probable errors of the means, the standard deviations, and the probable errors of the standard deviations are given in Table 4.

¹⁸ Holzinger, Karl J.: *Statistical Methods for Students in Education*. New York: Ginn, 1928. Pp. 372. (p. 243)

In considering this table the striking fact which appears is the similarity of the coefficients of variation in all cases. The probable errors of the coefficients vary from 3.3 to 4.4. Hypothetically, if we were to obtain the reliability of the differences between these coefficients of variation, if both probable errors were 3.0 the probable error of the difference would be about 4.2 and, therefore, the actual differences between the coefficients would have to be at least twelve units or more to be bordering on significance. There were no actual differences greater than nine. It would appear, therefore, that there are no reliable differences between the coefficients of variation. However, since a correlation between the mean and the standard deviation may be operating to cover up the differences in the coefficients of variation, it is necessary to go back to the reliability of the differences between the means and standard deviations to discover if there is any absolute difference in the dispersion of the groups compared. It is evident that in those cases where both the means and the standard deviations increased significantly that they were correlated and thus the differences in variability as indicated by the coefficient of variation were covered. The means and standard deviations were, however, in these cases increasing proportionately. In the other instances where the mean increased significantly but the standard deviation did not it would appear to be the case that something was operating to inhibit the dispersion of scores of these groups, but since we find no reliable differences between the coefficients of variation this factor is not significant. In general, it can be said that motivating factors are not significantly operating to inhibit the dispersion of scores on this test.

INFLUENCE OF INCENTIVES IN MAINTAINING EFFICIENCY

Since seven trials were given on any one testing day it was of interest to know whether the incentives given had any influence on efficiency during any one trial. In order to answer this question the percentage of subjects who decreased, remained the same, or increased their scores on the last trial over the second (the first trial was always omitted in the calculations) was computed. The results are given in Table 5.

The reliability of the differences of subjects by groups decreasing, remaining the same, or increasing their scores on the last trial over the second, according to Series I, II, and III, are given in Table 6.

The formula¹⁹ used to obtain the probable error of a percentage fre-

$$\text{quency was: P.E.}_{fp} = .6745 \sqrt{\frac{fp(100 - fp)}{N}}.$$

The facts brought out by this table indicate that there were, in general, no reliable differences between groups on the different series for scores decreasing, remaining the same, or increasing. How-

Table 4

Number of Children, Means, Standard Deviations, Coefficients of Variation, and the Probable Errors of Variability by Groups in Series I, II, and III on the Motivation-Dynamometer Test

Group	Children	Mean	Probable Error of Mean	Standard Deviation	Probable Error of Standard Deviation	Coeffi- cient of Vari- ation	Probable Error of Coeffi- cient of Variation
Series I							
Group A	50	6.480	.2922	3.060	.2066	47.22	4.4
Group B	57	6.395	.2475	2.770	.1750	43.32	3.6
Group C	53	6.896	.2784	3.005	.1969	43.58	3.8
Group D	51	6.598	.2607	2.760	.1843	41.83	3.6
Series II							
Group A	50	6.280	.279	2.928	.198	46.62	4.3
Group B	57	8.184	.308	3.445	.218	42.09	3.5
Group C	53	8.877	.351	3.783	.248	42.62	3.7
Group D	51	9.794	.364	3.847	.257	39.28	3.3
Series III							
Group A	50	6.220	.277	2.899	.196	46.61	4.3
Group B	57	8.991	.309	3.460	.219	38.48	3.1
Group C	53	10.462	.398	4.296	.281	41.06	3.5
Group D	51	10.225	.418	4.424	.395	43.27	3.8

ever, the differences were greatest between those who increased or decreased. It is apparent that the majority of subjects tended to decrease their scores in any seven trials. The next question which was raised was whether there was a difference in the percentage of subjects who decreased according to the type of incentive given. To ascertain this, the percentage of subjects by groups who decreased their scores 10 per cent or more, 15 per cent or more, and 25 per cent or more according to Series I, II, and III were then computed.

¹⁹ Holzinger, Karl J.: Statistical Methods for Students in Education. New York: Ginn, 1928. Pp. 372. (p. 243)

The percentage frequencies and the probable errors of the percentage frequencies are given in Table 7.

The reliability of the differences of the percentage of subjects by groups who decreased their scores by 10 per cent or more, 15 per cent or more, or 25 per cent or more according to Series I, Series I with Series II, Series I with Series III, and Series II with Series III is given in Table 8.

The facts which are brought out in this table indicate that there are no statistically significant differences between groups in Series I with the exception perhaps of Group A and D where, with a ratio difference of 2.92, there are ninety-seven chances in one hundred that the true difference is greater than zero. This difference approaches significance.

Table 5

Percentage of Scores on Subjects Decreasing, Remaining the Same, or Increasing and the Probable Errors of Percentage Frequency on the Last Trial over the Second Trial according to Groups A, B, C, D in Series I, II, and III on the Motivation-Dynamometer Test

Series	Decrease		Same		Increase	
	Score	P.E. fp	Score	P.E. fp Per Cent	Score	P.E. fp
Group A						
Series I	84.00	3.49	2.00	1.33	14.00	3.30
Series II	72.00	4.27	6.00	2.25	22.00	2.95
Series III	78.00	4.95	4.00	1.88	18.00	3.86
Group B						
Series I	84.16	3.25	1.75	1.16	14.01	3.08
Series II	82.42	3.39	10.52	2.73	7.01	2.27
Series III	75.38	3.83	3.50	1.63	21.03	3.62
Group C						
Series I	77.32	3.25	1.88	1.25	20.73	3.08
Series II	75.44	3.98	5.66	2.13	18.84	3.62
Series III	67.90	4.32	1.88	1.25	30.17	4.24
Group D						
Series I	82.32	3.60	1.96	1.30	15.68	3.43
Series II	66.64	4.45	1.96	1.30	31.36	4.37
Series III	60.77	4.60	5.88	2.21	33.32	4.45

Table 6
Reliability of Differences of Subjects Who Decreased, Remained the Same, or Increased Their Scores
on the Last Trial over the Second Trial according to Groups A, B, C, and D in Series
I, II, and III on the Motivation-Dynamometer Test

Series	Group A				Group B				Group C				Group D			
	Difference in Means	Probable Error of Difference	Ratio of Difference to Probable Error	Difference in Means	Probable Error of Difference	Ratio of Difference to Probable Error	Difference in Means	Probable Error of Difference	Ratio of Difference to Probable Error	Difference in Means	Probable Error of Difference	Ratio of Difference to Probable Error	Difference in Means	Probable Error of Difference	Ratio of Difference to Probable Error	Difference in Means
Decreased																
Series I and II	.12	5.51	2.17	2.02	4.69	.43	1.88	5.55	.33	15.68	5.72	2.74				
Series I and III	.60	5.26	1.14	8.78	5.02	1.74	9.42	5.79	1.62	17.64	5.75	3.74				
Series II and III	.60	5.81	1.03	7.04	5.11	1.37	7.54	5.84	1.11	5.87	6.40	.91				
Increased																
Series I and II	.80	5.14	1.55	7.00	3.82	1.83	1.89	5.21	.36	15.68	5.55	2.82				
Series I and III	.40	4.92	.81	7.02	4.75	1.47	9.44	5.65	1.67	17.64	5.61	3.14				
Series II and III	.40	5.38	.74	14.02	4.27	3.28	11.33	5.57	2.03	1.96	6.23	.31				
Same																
Series I and II	.40	2.61	1.53	8.77	2.96	2.96	3.78	2.46	1.53	3.92	0.00	0.00				
Series I and III	.20	2.30	.86	1.75	1.99	.87	0.00	0.00	0.00	3.92	2.56	1.53				
Series II and III	.20	2.93	.68	7.02	3.17	2.21	0.00	0.00	0.00	3.92	2.56	1.53				

Table 7

Percentage of Subjects by Groups Who Decreased Their Scores
10, 15, or 25 Per Cent or More on Series I, II, and
III on the Motivation-Dynamometer Test

Group	Series I		Series II		Series III	
	Percentage	Probable Error of Percentage Frequency	Percentage	Probable Error of Percentage Frequency	Percentage	Probable Error of Percentage Frequency
10 Per Cent or More Decrease						
Group A	54.00	4.75	52.00	4.76	52.00	4.76
Group B	64.37	4.26	36.82	4.30	42.00	4.40
Group C	62.23	4.48	42.37	4.57	30.17	4.25
Group D	72.52	4.22	29.40	4.30	23.52	4.01
15 Per Cent or More Decrease						
Group A	48.00	4.76	36.00	4.59	46.00	4.76
Group B	52.59	4.46	24.54	3.85	28.03	4.01
Group C	47.17	4.63	24.51	3.88	15.08	3.32
Group D	56.84	4.68	17.64	3.50	9.80	.89
25 Per Cent or More Decrease						
Group A	30.00	4.37	16.00	3.62	24.00	4.07
Group B	26.29	3.93	5.25	1.82	10.50	2.72
Group C	28.28	4.17	9.43	2.70	9.43	2.70
Group D	25.48	3.94	3.92	1.83	3.92	1.83

When the effects of control-motivation were considered (Group A in Series I, II, and III) there were no significant differences. Thus approximately 50 per cent of the subjects reduced their scores 10 per cent or more on the last over the second trial. When Group B, Group C, and Group D were compared in Series I and II there were, in general, significant differences in the percentage of subjects who decreased their scores. These differences favor Series II. In other words there was a smaller percentage of subjects who decreased their scores by 10 per cent or more when given success-repetition. When groups were compared on Series I with Series III the reliabilities of the differences in general indicated significance since the lowest ratio difference of 3.73 indicates that there are ninety-nine chances in one hundred that the true differences are greater than zero. Again these results indicate that a greater percentage de-

Table 8
Reliability of Differences of the Percentage of Subjects by Groups Who Decreased Their Scores
by 10, 15, or 25 Per cent or More from the Second to the Last Trials, Series I,
Series I and II, Series I and III, and Series II and III on the
Motivation-Dynamometer Test

Decrease														
Groups	10 Per Cent or More					15 Per Cent or More					25 Per Cent or More			
	Difference in Means	Probable Error of Difference	Ratio of Difference to Probable Error of Difference	Difference in Means	Probable Error of Difference	Ratio of Difference to Probable Error of Difference	Difference in Means	Probable Error of Difference	Ratio of Difference to Probable Error of Difference	Difference in Means	Probable Error of Difference	Ratio of Difference to Probable Error of Difference	Difference in Means	Probable Error of Difference
Series I														
Group A ₁ and Group B ₁	10.87	6.38	1.70	4.59	6.52	.70	3.71	5.87	.630					
Group A ₁ and Group C ₁	8.23	6.53	1.26	.83	6.64	.13	1.72	6.04	.280					
Group A ₁ and Group D ₁	18.52	6.35	2.92	8.84	6.67	1.13	4.52	5.88	.770					
Group B ₁ and Group C ₁	2.64	6.18	.43	5.42	6.43	.84	1.99	5.73	.347					
Group B ₁ and Group D ₁	7.65	6.00	1.28	4.25	6.46	.66	.81	5.56	.145					
Group C ₁ and Group D ₁	10.29	6.15	1.67	9.67	6.59	1.47	2.80	5.74	.487					
Series I and II														
Group A ₁ and Group A ₂	2.00	6.72	.30	12.00	6.61	1.82	14.00	5.67	2.469					
Group B ₁ and Group B ₂	28.05	6.05	4.63	18.05	5.89	4.76	21.04	4.33	4.859					
Group C ₁ and Group C ₂	19.86	6.40	3.10	22.66	6.04	3.75	18.85	4.97	3.792					
Group D ₁ and Group D ₂	43.12	6.02	7.16	39.20	5.84	6.71	21.56	4.34	4.967					

Series I and III

Group A ₁ and Group A ₃	2.00	6.72	.30	2.00	6.73	.30	6.00	5.97	1.005
Group B ₁ and Group B ₃	22.81	6.12	3.73	24.56	6.00	4.09	15.79	4.78	3.303
Group C ₁ and Group C ₃	32.06	6.17	5.20	32.09	5.70	5.63	18.85	4.97	3.792
Group D ₁ and Group D ₃	49.00	5.82	8.42	47.04	4.76	9.87	21.56	4.34	4.967

Series II and III

Group A ₂ and Group A ₃	0.00	6.73	0.000	10.00	6.61	1.51	8.00	5.45	1.467
Group B ₂ and Group B ₃	5.24	6.15	.852	3.49	5.56	.63	5.25	3.27	1.605
Group C ₂ and Group C ₃	12.20	6.24	1.955	9.43	5.20	1.81	0.00	3.82	0.000
Group D ₂ and Group D ₃	5.88	5.88	1.000	7.84	3.61	2.17	0.00	2.58	0.000

creased their scores under control-motivation by 10 per cent or more than under failure-repetition, failure-reproof, and failure-punishment. Comparison of Series II with Series III by groups showed no significant differences between the percentage who decreased under success or under failure.

Consideration of results indicating the percentage of subjects who decreased their scores 15 per cent or more followed the same general trend as those decreasing 10 per cent or more. When the groups were compared according to Series I with II and Series I with III the ratio differences varied from 3.75 to 9.87 or from ninety-nine chances in one hundred to absolute certainty that the true differences are greater than zero. In other words, as before, a greater percentage decreased their scores under control-motivation than decreased under success or failure. The reliability of the differences between Series II and III was not significant.

In comparing the percentage of subjects who decreased their scores 25 per cent or more the same general trend was again found. Between Series I and Series II there were ratio differences varying from 3.79 to 4.97 or from ninety-eight chances in one hundred to absolute certainty that the true differences are greater than zero. The ratio differences between Series I and Series III varied from 3.79 to 4.97 or from ninety-nine chances in one hundred to absolute certainty that the true differences are greater than zero. Between success and failure there are no significant differences.

INTERPRETATION OF RESULTS

From these results it may be concluded that there is a greater percentage of subjects who decrease their scores from second to last trial than increase in any six trials given at one time. In addition, a greater percentage decreased their scores by 10, 15, or 25 per cent when given control-motivation, than under success-repetition, success-praise, success-reward, failure-repetition, failure-reproof, or failure-punishment even though under each of these latter types of motivation the groups as a whole increased their energy out-put. It may also be concluded that these types of motivation not only increase efficiency but tend to more nearly maintain initial efficiency, since a lesser percentage decreased their efforts by 10, 15, or 25 per cent or more in any six trials given at one time than did those given control-motivation.

QUALITATIVE ASPECTS OF THE INFLUENCE OF MOTIVATION

When the method and procedure of the study were being worked out during preliminary tests certain qualitative aspects of the influence of motivation were noted. These items were incorporated in the series of record blanks (Figures 6 and 7) and include such factors as expressive behavior, verbal comments, attention, and mood. Since the experimenter's attention was entirely taken up by the demands of the test situation itself, it was necessary to have the presence or degree of presence of these items checked by a recorder.

RELIABILITY OF RECORDERS

During the study it was necessary to use two different recorders, one (X) who checked items for all subjects in Iowa City and another (Y) who checked items for subjects in Ames. In order to have some check on the reliability of the recorders X and Y, two other persons S and R were asked to come into the test situation and check simultaneously with X or Y. It was not possible to obtain an index of the reliability of X with Y, however, since the two individuals were in different cities and the experimenter was not able to give adequate attention for such a check. The results on the reliability of recorders were as follows: recorders X and S each checked a total of 246 items on twenty-four subjects. The reliability coefficient was $.868 \pm .011$. Recorders Y and R checked a total of 285 items on twenty-five subjects. The coefficient was $.755 \pm .017$. Recorder S checked items with recorder X during an early part of the experiment when neither X nor S had experience in recording. Recorder R, however, was not able to give any time for recording until the experiment had progressed for five weeks. Therefore, recorder Y had the advantage of several weeks training or experience which R had not had. This may account in part for the lower correlation between Y and R than between X and S. In checking of this kind, it is preferable to ascertain the reliability of recorders at the beginning of the testing period and continue practice until it is .90 or more, if necessary. The experimenter was unable to do this in Ames because no one was available who could give time for such checking. It should also be noted that the reliability coefficients represent approximately five minutes of more or less intermittent observation on each subject.

EXPRESSIVE BEHAVIOR

The items under the heading expressive behavior will be found in

the series of record blanks (Figures 6 and 7). The presence or degree of presence of all items during each trial was checked in the appropriate space on this blank. Item 1 (contorts and tenses face and neck) and Item 7 (smiles happily, elated) under this category were rated on a basis of three points if present. In these cases the maximum score which a subject could obtain in seven trials would be twenty-one. In all other items the maximum score for any seven trials would be seven.

The frequencies with which the various items were checked according to groups and Series I, II, and III were tabulated. In general, there seemed to be a tendency for items on Series II and III to be checked more frequently than those in Series I, and for the first seven and last items to be most frequently checked. Accordingly, the reliability of the differences of these latter items was computed. Since Item 1 (contorts and tenses face and neck) and Item 7 (smiles happily, elated) were rated on a basis of three points these were treated separately.

The frequencies of Item 2 (breathes hard, face reddens), Item 3 (lifts elbow, arm), Item 4 (uses shoulder and back muscles), Item 5 (attempts to use two hands), Item 6 (takes extra squeeze), and Item 16 (no change) were translated into relative frequencies of occurrence (n) and based on the possible number of times they could be checked. The probable errors of these relative frequencies

were obtained by the formula:²⁰ $P.E._{(p)} = .6745 \sqrt{\frac{p' q'}{n}}$

In general, for Group A there was a significant tendency for Item 2 (breathes hard, face reddens), Item 5 (attempts to use two hands), and Item 6 (takes extra squeeze) to decrease in frequency checked, while Item 16 (no change) increased in frequency checked from Series I to Series II and Series I to Series III. Item 3 (lifts elbow, arm) and Item 4 (uses shoulder and back muscles) were checked more frequently and significantly in Series II and III for this group. With these two exceptions there seemed to be a tendency for these expressive behavior items to be checked less frequently, and for Item 16 (no change) to be checked more frequently in Series II and Series III than in Series I.

In Group B there was a definite and significant tendency for all items to be checked more frequently under Series II and Series III

²⁰ Rietz, Henry Lewis [editor]: *Handbook of Mathematical Statistics*. Boston: Houghton, Mifflin. [c. 1924] Pp. 221. (p. 77)

than in Series I, and in Series III as compared with Series II. This same trend is found in Group C and Group D.

As to Item 1 (contorts and tenses face and neck) the mean of Group A decreases slightly but not significantly so in Series II, although the decrease approaches significance when Series I and III are compared. In Group B this item is more frequently checked in Series II and III than in Series I, and also in Series III than in Series II.

This same tendency is found in Group C and D. There is then, in general, a significant tendency for Item 1 to be checked less frequently in the control-motivated series and more frequently in the motivated series.

Item 7 (smiles happily, elated) is checked less frequently for Group A in Series II as compared with Series I (this difference approaches significance), less frequently in Series III as compared with Series I (this difference is statistically reliable), and less frequently in Series III than in Series II (difference approaches significance). (See footnote page 86.) In Groups B, C, and D there was a significant increase in frequency with which this item is checked, while in Series II as compared with Series I and in Series I as compared with Series III the decrease over Series I is again significant. Likewise the decrease in Series III over Series II was notable. There seemed to be a significant tendency to smile happily under success-repetition, success-praise, and success-reward and not to smile happily under failure-repetition, failure-reproof, and failure-punishment.

It would seem to be the case then that one influence of motivation is a tendency to increase the amount of energy output that is directed through other channels (neck, face, shoulders, etc.) than through the main channel (the hand and arm).

SUBJECTS LIKING GAME I OR II

Shortly after the second week of the experimental procedure had begun, it occurred to the experimenter that it would be of interest to ask the subject which game he had liked best at the close of the test period for that day. The tests were alternated with every other subject in order to avoid either one or the other being chosen because it was the first or last. In Series I there was a definite tendency for a decidedly greater number of subjects to choose the Perforation Test (Game II) as the game they preferred best, rather than Presentation I on the motivation-dynamometer. Under Series

II a greater percentage of the control group still preferred Game II to the first game. In Groups B, C, and D, however, a greater percentage now chose Presentation 2 on the motivation-dynamometer rather than the Perforation Test. In Series III a greater percentage still preferred the Perforation Test to the first test in Group A. A greater percentage of those in Group B and D preferred the motivation-dynamometer (Presentation 3) even though they failed to attain their goal. A greater percentage in Group C, however, preferred the Perforation Test to being reproved on the motivation-dynamometer at which they were failing.

In every case on Series I there were decidedly significant differences in favor of those in all groups who preferred the Perforation Test to Presentation 1 on the Motivation-Dynamometer Test. In Series II three differences are statistically significant and one approaches significance. Group A still preferred the Perforation Test to Presentation 1 on the first game. Groups C and D preferred Presentation 2 on the first game, while Group B probably preferred it.

On Series III all differences are significant; Group A still preferred the Perforation Test while Groups B and D preferred Presentation 3 on the first game even though they failed to achieve their goal and in one case were punished for not doing so. Group C preferred the Perforation Test to being reproved because they did not attain the goal on Presentation 3 of the first test.

In giving the Perforation Test the experimenter tried not to give any added incentive. No mention of subject's performance was made either of how well or how poorly he was doing. In many respects the external motivation conditions of this test were similar to those given control-motivation in the motivation-dynamometer series except that the subject in this case did have visual knowledge of both progress and results (punching out the holes).

In general, it may be stated that in this study a significantly larger percentage of subjects preferred a game with some motivation to one that was given under control-motivation even though they were not successful in attaining the goal set.

ATTENTION

Since some investigators have considered that the attention factor was important in motivation this item was included on the record blank and checked both during and after the trial. If a subject's eyes were focused "on" the game or apparatus during the trial the experimenter put a "1" in the appropriate space. If the eyes were

not focused on the apparatus the attention was marked as "off." Part of the time both were checked to indicate that attention wandered. The duration of any one trial was about a minute in length. The same procedure was used in checking attention "after" the trial. The maximum score which could be obtained for any one of the categories was seven.

In general, during trial there was a decided tendency for attention to be on game rather than off game. In Group A there was a slight tendency for attention to be checked more frequently both on and off game (i.e., wanders) in Series III than in Series I or II. There was also a tendency for attention to be checked less frequently on game in Series III than in Series I or II and conversely checked more frequently as off game in Series II and III than in Series I. In Groups B, C, and D there was a tendency for attention to be checked more frequently on game in Series II and Series III than in Series I and for attention off game to be checked less frequently in Series II and III as compared with Series I.

There was a decided tendency during trial for attention to be focused on the game regardless of series or groups considered. After trial, there is probably a significant tendency for attention to be off game more than on game when control-motivation (Group A) is given. When Groups B, C, and D are given Series II or Series III there is a significant tendency for attention to be on the game rather than off the game.

When frequency of occurrence of attention on game, off game, both on and off game during trial is compared there is a tendency, which is probably significant, for attention to decrease on game during trial and to increase off game for Group A in all series and for Groups B, C, and D in Series I. Groups B, C, and D increased significantly in frequency with which attention was on game during trial, and also decreased significantly for off game during trial and also for both on and off game in Series I over Series II, while there was a decrease in frequency of attention on game for these groups in Series I over Series III, and an increase in both on and off game for Series II over Series III.

In general, these data tended to indicate that attention was more frequently checked on the game during a trial for a motivated series and less frequently for the control-motivated series. In the latter case the tendency to decrease attention on game and to increase off game was significant. Although attention was more on game than

off in Series III than in Series I and Series II, it is less in Series III than in Series II and this decrease is significant.

After trial there is a slight tendency for attention to be checked more frequently off game than on game for all groups in Series I, and for Group A on Series II with III and Series I with III. During the motivated series, attention after trial is significantly more on game than off game. Also both on and off game is more frequently checked than in Series I and II. However, although in Series III, Groups B, C, and D in general focused attention on game rather than off game, there is a general tendency for it to be significantly less so on Series III as compared with Series II.

In Series III attention both on and off game is significantly increased over both Series I and II.

In general, attention during game and after game is focused more on game than off game in Groups B, C, and D when motivation is given. There is a significant tendency for attention to be off game both during and after trial in the control-motivation series. In addition, although attention is on game more than off game both during trial and after trial in Series III (failure), it is less so than in Series II (success).

Mood

There were three ratings on mood for each subject during each testing period: (1) mood as evidenced from entrance until given first directions; (2) mood during Game I; (3) mood as evidenced from end of last trial in Game I until beginning of Game II. The recorders were instructed to check the mood of the subject at any point along the line, elated-depressed. In assigning a numerical value to these ratings the line was divided into six equal parts. One-fourth of an inch on either side of normal was given a value of three with a value of zero at the elated end and a value of six at the depressed end.

There were no significant differences in the ratings of Mood 1 according to Series I with II, Series I with III, or Series II with III. There are no differences in Mood 2 as rated for Group A. When rating Group C and D between Series I and II there is a significant tendency for the ratings on Series II to be more toward the elated end. When Series I is compared with Series III (failure) for Groups B, C, and D the ratings tend to be significantly more toward the depressed end. The differences between Series II and III are significant for Mood 2.

Ratings for Mood 3 indicate that there is a slight tendency (but only approaching significance) for Mood 3 to be rated more toward the elated end in Groups B, C, and D under Series II and more toward the depressed end in Series III. Only for Group C in the latter case, however, does this tendency even approach significance.

In general, then during success there is a significant tendency to be elated, while under failure the tendency is to be depressed as rated in this study.

AGE AND SEX DIFFERENCES

In order to obtain some indication as to the influence of these different types of incentives at the various age levels, the percentages of two-, three-, four-, five-, six-, and seven-year-olds increasing their score 30 per cent or more, 50 per cent or more, or 100 per cent or more in Series II over Series I, Series III over Series I, and Series III over Series II were computed. These percentages were determined and, in general, showed that there was a tendency for a greater percentage of younger subjects to increase their scores in Series II and III over Series I than for the older subjects. The reliabilities of the differences approached significance and in some instances were above four probable errors. Just what factors are operating to inhibit the initial performance of the younger children is not clear. It may be that the younger child needs more motivation in comparison with the older child for effective performance than that given in the control-motivation series. A more probable factor is that physiological age or conditions are in some way accounting for the lower performance on this initial test. An older child may be able to adjust his hand and arm muscles to the apparatus much more effectively than the younger child because of better coördination and thus better symmetry of action. If age norms were available it might help solve the problem as to whether the performance of the younger group on the control-motivation series differs from the average performance for that age level. However, chronological age norms may not be the optimum criterion since physiological age seems to be a better index of muscular development. Unfortunately, neither type of age norm is available for these groups.

In order to ascertain whether or not there were sex differences in response to the types of motivation given, the percentages of boys and girls who increased 30 per cent, 50 per cent, 100 per cent or more on each series were computed. The standard for comparison

Table 9

Percentage of Children Who Increased Their Scores by 30, 50, or 100 Per Cent or More according to Sex-Group and Test-Group on Series II over Series I, Series III over Series I, and Series III over Series II on the Motivation-Dynamometer Test

Groups Compared	30 Per Cent or More		50 Per Cent or More		100 Per Cent or More	
	Percentage	Probable Error of Percentage Frequency	Percentage	Probable Error of Percentage Frequency	Percentage	Probable Error of Percentage Frequency
Series II over Series I						
Boys						
Group A	4.00	1.86	0.00	0.00	0.00	0.00
Group B	48.88	4.45	37.50	4.29	3.13	1.54
Group C	40.74	4.55	11.11	2.96	0.00	0.00
Group D	71.40	4.26	57.12	4.61	24.99	4.08
Girls						
Group A	4.00	1.86	4.00	1.86	0.00	0.00
Group B	48.00	4.40	20.00	3.56	0.00	0.00
Group C	53.84	4.99	46.15	4.60	11.54	2.94
Group D	78.21	3.86	47.79	4.70	4.34	1.91
Series III over Series I						
Boys						
Group A	4.00	1.86	4.00	1.86	15.63	3.28
Group B	59.38	4.37	43.75	4.41	15.63	3.28
Group C	59.25	4.54	51.85	4.60	7.41	2.41
Group D	17.85	3.54	14.28	3.29	3.57	1.74
Girls						
Group A						
Group B	60.00	4.37	4.80	4.45	16.00	3.26
Group C	76.92	3.87	53.84	4.99	26.92	4.09
Group D	17.37	3.57	4.34	4.34	0.00	0.00
Series III over Series II						
Boys						
Group A	4.00	1.86	0.00	0.00	0.00	0.00
Group B	18.75	3.47	12.50	2.95	3.70	0.00
Group C	22.22	3.82	11.11	2.96	10.71	0.00
Group D	71.42	4.26	46.41	4.70	2.91	0.00
Girls						
Group A	4.00	1.86	0.00	0.00	0.00	0.00
Group B	24.00	3.80	12.00	2.89	0.00	0.00
Group C	23.08	3.96	19.23	3.64	3.85	1.77
Group D	47.77	4.70	21.71	3.24	4.34	1.91

Table 10
Reliability of Differences between Boys and Girls Who Increased Their Scores by 30, 50, or 100 Per Cent or More according to Series I and II, Series I and III, and Series II and III on the Motivation-Dynamometer Test

Increase												
30 Per Cent or More			50 Per Cent or More				100 Per Cent or More					
Groups		Boys	Girls	Difference in Means	Probable Error of Difference	Series I and II						
Ratio of Difference to Probable Error of Difference	Difference to Probable Error of Difference					Difference in Means	Probable Error of Difference	Ratio of Difference to Probable Error of Difference	Difference to Probable Error of Difference	Difference in Means	Probable Error of Difference	Ratio of Difference to Probable Error of Difference
Group A and Group A				0.00	1.86	0.00	+ 4.00	1.86	2.15	0.00	0.00	0.00
Group B and Group B				+ 1.13	6.25	.18	+17.50	5.57	3.14	+ 3.12	1.54	2.02
Group C and Group C				-13.11	6.75	1.94	-35.04	5.46	6.41	-11.53	2.94	3.92
Group D and Group D				- 6.81	5.74	1.18	+ 9.33	6.58	1.41	+20.65	4.50	4.58
Series I and III												
Group A and Group A				+ 4.00	1.86	2.15	+ 4.00	1.86	2.15	0.00	0.00	0.00
Group B and Group B				- .63	6.17	.10	- 4.25	6.26	.67	- .38	4.62	.08
Group C and Group C				-17.67	5.96	2.96	- 2.00	6.50	.31	-19.52	4.74	4.11
Group D and Group D				+ .48	5.02	.09	+ 9.94	3.80	2.61	+ 3.75	1.74	2.05
Series II and III												
Group A and Group A				0.00	1.86	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Group B and Group B				- 5.25	5.14	1.02	+ .50	4.12	.12	- .14	2.48	.05
Group C and Group C				- .85	5.50	.15	- 8.12	4.69	1.73	+ 6.37	3.47	1.86
Group D and Group D				+23.65	6.34	3.73	+24.70	5.70	4.33	0.00	0.00	0.00

A minus indicates that the greater percentage is in favor of the girls, a plus that it is in favor of the boys.

was Series I and Series II. The percentages are given in Table 9. The reliabilities of these differences are given in Table 10. In general, the reliabilities of the differences are not significant although they approach significance in some instances but no outstanding differences are apparent. The differences which do exist are probably due to individual differences which were computed on a percentage basis; actual differences may, however, exist.

Because of the limited number of cases in each test group it is not possible to draw any conclusions as to the relationship between either age or sex and motivation as given in this study.

RELIABILITY OF THE MOTIVATION-DYNAMOMETER TEST

The reliability of the Motivation-Dynamometer Test was computed in two ways: first, by correlating the average scores on Series I, II, and III made by the fifty subjects in Group A, which acted as the control group; second, by correlating the sum of the scores on the odd trials with the sum of the scores on the even trials on Series I for 211 subjects ranging in age from twenty-seven months to ninety-six months. In order to take into account the influence of the wide age range used in computing this last coefficient, the data were also divided into three large age groups.

The reliability coefficients obtained by correlating the average scores on Series I, II, and III, made by the fifty children in Group A ranging in age from twenty-eight months to ninety-two months, together with the coefficients of alienation and the indices of reliability were:

Series	r	P.E.	Coefficient of Alienation	Index of Reliability
Series I and II	.929 $\pm$.013		.371	.963
Series I and III	.9003 $\pm$.018		.435	.948
Series II and III	.939 $\pm$.011		.344	.969

The reliability coefficients obtained by correlating the sum of the odd and even trials and then stepping up by the Spearman-Brown formula on Series I for subjects ranging in age from twenty-seven to fifty-nine months, from sixty to seventy-one months, from seventy-two to ninety-six months, and from twenty-seven to ninety-six months, together with the coefficients of alienation and the indices of reliability were:

Children	Reliability Between Chance Halves	Coeffi- cient of Relia- bility	Coeffi- cient of Alienation	Index of Relia- bility
57	.967 $\pm$.005	.983	.256	.980
86	.881 $\pm$.016	.937	.475	.940
68	.957 $\pm$.007	.978	.291	.978
211	.975 $\pm$.002	.987	.222	.987

INTERPRETATION OF RESULTS

In general, the reliability coefficients obtained between two applications of the test, one week apart, were somewhat less than the self-correlation obtained from a single application of the test.

Both small numbers and wide age range were also obviously affecting the correlations. The coefficients of alienation, in general, ran rather high. For example, the coefficient of alienation of .475 computed using the reliability coefficient for the odd and even trials given on Series I for the sixty to seventy-one month age group indicates that one is able to reduce the error of his estimates only a little less than one-half from what would be the case if he were making a pure guess. Interpreted in this light, the reliability coefficient is not exceptionally high, especially if one were to consider individual differences. It is, however, probably high enough to distinguish group differences.

INFLUENCE OF INCENTIVES UPON PERFORMANCE IN PERFORATION TEST

The purpose of giving the Perforation Test immediately after the Motivation-Dynamometer Test was to ascertain whether or not the facilitating or inhibiting effect of any of the various types of incentives on the Motivation-Dynamometer Test would be carried over to another test which immediately followed as indicated by number of holes punched.

The means, the probable errors of the means, the standard deviations, and the probable errors of the standard deviations of the scores (holes punched) of subjects by groups on Series I, II, and III of the Perforation Test are given in Table 11.

The reliability of the differences between the means and standard deviations of scores on the number of holes punched of subjects by groups on Series I, II, and III, and Series I with Series II, Series I with Series III, and Series II with Series III on the Perforation Test are given in Table 12.

Table 11

Number of Children, Means, Probable Errors of the Means, Standard Deviations, and Probable Errors of the Standard Deviations of Subjects by Groups on Series I, II, and III on the Perforation Test

Group	Children	Mean	Probable Error of Mean	Standard Deviation	Probable Error of Standard Deviation
Series I					
Group A ₁	50	40.30	1.459	15.30	1.032
Group B ₁	57	41.27	1.353	15.15	.957
Group C ₁	55	40.24	1.239	12.76	1.182
Group D ₁	51	39.85	1.266	13.40	1.266
Series II					
Group A ₂	50	48.30	1.798	18.85	1.271
Group B ₂	57	49.08	1.648	18.45	1.165
Group C ₂	53	47.97	1.436	15.50	1.016
Group D ₂	51	48.97	1.629	16.45	1.099
Series III					
Group A ₃	50	51.00	1.936	20.30	1.369
Group B ₃	57	52.68	1.862	20.84	1.317
Group C ₃	53	52.22	1.691	18.25	1.196
Group D ₃	51	50.93	1.747	18.50	1.236

There were no statistically significant differences between the means of each of the groups in Series I since the highest ratio difference is .77, which indicates that there are only approximately seventy-one chances in one hundred that the true differences are greater than zero. The ratio differences as compared to the standard deviations of these groups run somewhat higher, although there are no differences greater than 1.91, indicating that the greatest difference is not statistically significant since there are only eighty-nine chances in one hundred that the true difference is greater than zero. On the whole we may conclude that probably these groups are comparable as judged by their initial performance on the Perforation Test.

The ratio differences between the means of each of the groups in Series II indicate that none of the differences are statistically re-

Table 12

Reliability of the Differences between the Means and Standard Deviations of Scores on Number of Holes Punched by Groups on Series I, II, and III, and Series I and II, Series I and III, and Series II and III on the Perforation Test

Groups	Scores on Number of Holes Punched					
	Difference in Means	Probable Error of Difference	Ratio of Difference to Probable Error of Difference	Difference in Means	Probable Error of Difference	Ratio of Difference to Probable Error of Difference
Series I						
Group A ₁ and Group B ₂	— .97	1.99	— .49	.15	1.407	.11
Group A ₁ and Group C ₂	.06	1.88	.03	2.54	1.328	1.91
Group A ₁ and Group D ₂	.45	1.93	.23	1.90	1.366	1.39
Group B ₁ and Group C ₂	1.03	1.30	.57	2.39	1.271	1.88
Group C ₁ and Group D ₂	.39	1.73	.23	— .64	1.223	1.34
Group B ₁ and Group D ₂	1.42	1.85	.77	1.75	1.310	1.34
Series II						
Group A ₂ and Group B ₂	— .78	2.44	— .32	.40	1.724	.23
Group A ₂ and Group C ₂	.33	2.30	.14	3.35	1.627	2.06
Group A ₂ and Group D ₂	— .67	2.38	— .28	2.40	1.680	1.43
Group B ₂ and Group C ₂	1.11	2.19	.51	2.95	1.546	1.91
Group B ₂ and Group D ₂	.11	2.27	.05	2.00	1.602	1.25
Group C ₂ and Group D ₂	— 1.00	2.12	— .47	— .95	1.497	— .63
Series III						
Group A ₃ and Group B ₃	.168	2.69	.62	.54	1.900	.28
Group A ₃ and Group C ₃	.122	2.57	.47	2.05	1.818	1.13
Group A ₃ and Group D ₃	.070	2.61	.03	1.80	1.844	.98
Group B ₃ and Group C ₃	.460	2.52	.18	2.59	1.779	1.46
Group B ₃ and Group D ₃	1.750	2.55	.69	2.34	1.806	1.30
Group C ₃ and Group D ₃	1.290	2.43	.53	.25	1.720	.15
Series I and II						
Group A ₁ and Group A ₂	8.00	.840	9.50	3.55	.595	6.00
Group B ₁ and Group B ₂	7.81	.732	10.80	3.30	.498	6.60
Group C ₁ and Group C ₂	7.73	.789	9.80	2.74	.559	4.90
Group D ₁ and Group D ₂	9.12	.798	11.50	3.05	.563	5.40
Series I and III						
Group A ₁ and Group A ₃	10.70	.938	10.40	.50	.663	7.50
Group B ₁ and Group B ₃	11.41	.992	11.40	.50	.663	7.50
Group C ₁ and Group C ₃	11.98	.985	12.20	5.49	.697	7.90
Group D ₁ and Group D ₃	11.80	.982	11.30	5.10	.695	7.30
Series II and III						
Group A ₂ and Group A ₃	2.70	.837	3.20	1.45	.592	8.40
Group B ₂ and Group B ₃	3.60	.733	4.90	1.45	.592	2.40
Group C ₂ and Group C ₃	4.25	.636	6.70	2.75	.449	6.10
Group D ₂ and Group D ₃	1.96	.635	3.10	2.05	.450	4.60

liable since the highest ratio of the actual difference to the probable error of the difference is .51, indicating approximately sixty-four chances in one hundred that the true difference is greater than zero. Again, the ratio differences between the standard deviations of these groups are consistently somewhat higher than those between the means. The highest difference is 2.06, which shows that there are ninety-one chances in one hundred that the true difference is greater than zero. However, since there were eighty-nine chances in one hundred that the difference between Group A and Group C was greater than zero to begin with (see Series I), it is probable that the difference of 2.06 is not due to motivation. On the whole, therefore, the results indicate that it is probable there are no statistically significant differences between either the means or the standard deviations of these groups.

Again, the ratios of the differences to the probable errors of the differences for all means according to groups in Series III were not statistically significant since the highest ratio of .62 indicated that there were only sixty-seven chances in one hundred that the true differences were greater than zero. Again, the ratio differences between the standard deviations have a tendency to run slightly higher than those between the means, but they are still insignificant since the highest ratio, 1.46, shows that there are only eighty-four chances in one hundred that the true differences are greater than zero.

The intercorrelations in terms of Pearson product moment correlation between the test Series I, II, and III for each of the four groups, A, B, C, and D, on the Perforation Test were as follows:

Series	Children	r	P.E.
Series I and II	50	.887 ± .019	
Series I and III	50	.884 ± .022	
Series II and III	50	.902 ± .018	
Series I and II	57	.908 ± .015	
Series I and III	57	.864 ± .023	
Series II and III	57	.920 ± .014	
Series I and II	53	.835 ± .027	
Series I and III	53	.822 ± .031	
Series II and III	53	.930 ± .013	
Series I and II	51	.860 ± .025	
Series I and III	51	.854 ± .026	
Series II and III	51	.933 ± .013	

INTERPRETATION OF RESULTS

From this tabulation it is evident that the differences between all means are probably statistically significant, since all except two of them are above four probable errors which indicates absolute certainty and only two are below four, one being 3.1 and one being 3.2, which still indicates that there are ninety-eight chances in one hundred that the true differences are greater than zero.

It may be concluded then that although there were significant differences between test series (practice effects), there seems to be very little indication that there are carry-over effects due to the different types of incentives used in a previous task.

RELIABILITY OF THE PERFORATION TEST

Although the Perforation Test has been used in other studies and has been considered a fairly reliable test it was thought necessary to include some indication of the reliability of the test as given under the slightly different conditions in this study, namely of presenting three trials all on the same day rather than one trial on three successive days. The reliability was computed in two ways: (1) the average of each of the three trials given on one day was correlated for Series I, II, and III for subjects who were in the control group; (2) the sum of the holes punched for the tenth, thirtieth, and fiftieth second were correlated with the sum of the holes punched for the twentieth, fortieth, and sixtieth second for all subjects given Series I.

The reliability coefficients obtained by correlating the scores of fifty subjects in the control group, ranging in age from twenty-eight months to ninety-two months, given two applications of the test under the same conditions were:

Groups	r	P.E.	Coefficient of Alienation	Index of Reliability
Group A and Group A	.887 $\pm$.020		.462	.940
Group A and Group A	.884 $\pm$.020		.467	.940
Group A and Group A	.902 $\pm$.017		.432	.949

The reliability coefficients computed by the second method according to Trials 1, 2, and 3 and then stepped up by the Spearman-Brown formula on Series I for 211 subjects ranging in age from twenty-seven months to ninety-six months, together with the coefficients of alienation and the indices of reliability were:

Trial	Reliability Between Chance Halves	Coeffi- cient of Reliability	Coeffi- cient of Alienation	Index of Relia- bility
Trial 1	.92 $\pm$.007	.960	.384	.961
Trial 2	.94 $\pm$.005	.970	.332	.971
Trial 3	.94 $\pm$.006	.967	.353	.967

INTERPRETATION OF RESULTS

The reliability coefficients computed by the second method are all higher than those calculated by taking scores made on two presentations of the test given one week apart. The coefficients of alienation indicate that we were only able to reduce the error of estimate about one-third.

The consistency with which this test measures what it purports to measure is high enough for use in this study to indicate group differences. There are, at present, no other studies which have published reliability coefficients on this same test, so that it is not possible to compare it with those obtained by other investigators.

CHAPTER VI

SUMMARY, RESULTS, AND CONCLUSIONS

The importance and significance of motivation have been pointed out more generally as a basis of explanation in child psychology and more specifically as it concerns at least three phases of development in the life of the young child: (1) as a means of social control; (2) in the formation of habits and in the acquisition of skills and knowledge; and (3) in building up a satisfactory attitude toward the facing of reality.

A survey of the characteristic studies in the field indicated that very few experimental studies had been done on incentives with children below eight years of age. A number of problems were, therefore, raised with respect to the influence of external incentives upon the performance of tasks done by young children. These problems are, of course, too numerous to include in one study.

The present investigation represents, therefore, but one approach to certain aspects of the problem of motivation and its effectiveness for the performance of a task. The emphasis has been upon the external types of incentives rather than upon the more internal types of motivation designated as drives, urges, needs, and wants. The purpose of this investigation has been to obtain data as to the influence and relative effectiveness of the following kinds of incentives upon the performance of a task: control-motivation, success-repetition, success-praise, success-reward, failure-repetition, failure-reproof, failure-punishment.

The subjects who completed all three test-presentations of the experiment were 213 children who were obtained from the following sources: the preschool laboratories of the Iowa Child Welfare Research Station; Grades 1 and 2 of the University Elementary School of the State University of Iowa; the Nursery School at Iowa State College of Agriculture and Mechanic Arts; and from four kindergartens of the Ames public schools, Ames, Iowa. The age range of this group varied from twenty-seven months to ninety-six months, with ages taken from the date of their first test.

In order to be able to compare the relative effectiveness of any of these incentives with any other it was necessary to use at least four

comparable groups, since to give so many variables to each subject would lead to confusing results because of practice, fatigue, and carry-over effects. It was impossible to equate the groups on the basis of initial performance on Series I of the first test before giving any other variables. It was, therefore, necessary to assign subjects to a group on a random basis by listing the subjects in their school groups and numbering them accordingly. A pack of cards was numbered, shuffled, and dealt into four piles constituting Groups A, B, C, and D. A subject was, therefore, automatically put into the group into which his number fell. Each subject was given three test series, each approximately one week apart. Each time a subject was tested he was requested to perform two tasks, the Motivation-Dynamometer Test devised by the investigator and the Perforation Test which is used in the preschool psychological laboratories of the Iowa Child Welfare Research Station. Series I was the control-preliminary series given under Presentation 1 with the motivation-dynamometer. On Series II, Group A acted as the control group and was given control-motivation; Group B was given success-repetition; Group C was given success-praise; and Group D was given success-reward. On Series III, Group A again received control-motivation; Group B was given failure-repetition; Group C received failure-reproof; and Group D was given failure-punishment on Presentation 3 of the motivation-dynamometer. The effectiveness of the different types of motivation was measured in terms of the amount of energy expended during a squeeze on the motivation-dynamometer. Following the seven trials on this apparatus three trials each one minute in length were given on the Perforation Test. In addition to this quantitative measure of the influence of incentives upon performance, other more qualitative aspects were checked by a recorder who was present during the test situation and who checked the presence, or degree of presence, of certain items.

The following results were obtained:

1. By omitting two extreme cases from Group C, the means and standard deviations of all four groups on Series I given on the Motivation-Dynamometer Test were made comparable within 1.34 probable error units (82 chances in 100 that the true differences were greater than zero). The number of subjects, the average age, and the age range of the four groups were fairly similar.

2. The following findings deal with the influence of incentives given in Series II:

- (a) The means of the success-repetition, the success-praise, and the success-reward groups were all higher than the mean of the control-motivation group which decreased slightly but not significantly so over the initial test. The reliabilities of all differences were statistically significant, since the ratio of the differences to the probable errors of the differences was four or more.
- (b) The mean of the success-praise group was higher than the mean of the success-repetition group but the difference was not statistically significant. However, although the reliabilities of the differences between the means were not statistically significant, the chances of obtaining a better performance on the Motivation-Dynamometer Test by praising a child are greater than if one merely asked him to repeat his performance. One would be safer in using praise as given in this study than in not using it if effective performance were desired.
- (c) The mean of the reward group is higher than that of the praise group but the difference is not statistically reliable. However, the chances of obtaining a better performance by rewarding the child are greater than those of praising him as given in this study.
- (d) The mean of the reward group is higher than that of the repetition group and the reliability of the difference approaches significance. It is highly probable that success-reward is more effective than success-repetition for the performance of this task.

In summarizing the effectiveness of the types of incentives given during Series II it may be concluded that some motivation is more effective than the control-motivation for the performance of a task. It is highly important for effective performance that a subject be given at least visual knowledge of progress and results of his performance, together with an aim or goal to work toward.

It is probable that some motivation in addition to success-repetition is more effective for the performance of a task than merely asking the subject to repeat his performance when he has been successful in attaining his goal.

While the differences between the success-reward and the success-praise groups are not statistically reliable the chances of obtaining a better performance under success-reward are greater than those

under success-praise. It is probable, therefore, that success-reward as given in this experiment is more effective for performance of this task than success-praise.

3. The influence of incentives given in Series III was:

- (a) The means and standard deviations of the failure-repetition, the failure-reproof, and the failure-punishment group were all higher than the mean of the control-motivation group. The ratio differences between these groups were all statistically reliable.
- (b) The mean of the failure-reproof group is higher than that of the failure-repetition group. The reliability of the difference approaches significance.
- (c) The mean of the failure-punishment group is higher than that of the failure-repetition group. The reliability of the differences approaches significance.
- (d) The mean of the failure-reproof group is higher than that of the failure-punishment but the difference is not statistically significant.

In summarizing the results on the effectiveness of the incentives given on Series III it may be concluded that some motivation is more effective than control-motivation for the performance of a task, even though the subject fails to attain the goal set and, in addition, may be reproofed or punished because of it. It is probable that some motivation in addition to failure-repetition is more effective for performance than merely asking the subject to repeat the task. Reproof and punishment as given in this study are equally effective for the performance of the task. In evaluating the results on this series it is necessary to point out that failure has been preceded by success with the experimenter and that the train system may be more interesting to the young child than the water system.

4. The means of Groups B, C, and D on Series II were higher than for those under the control-motivation series, except the mean of Group A. The slight decrease in the mean of this group is not reliable. The increase in the means on Series II for Groups B, C, and D was statistically significant. These results indicate that some motivation is more effective for the performance of this task than control-motivation.

5. The means on Series III for Groups B, C, and D are all higher than those for Series I. The mean of Group A decreased on this series. The decrease in the case of the latter group was not

statistically significant, however. The differences between series for the other three groups were all statistically significant. It is concluded that some motivation is more effective for the performance of a task than is control-motivation, even though the subject is failing to attain the goal set and may, in addition, be reproved or punished for not doing so.

6. The means of Groups B, C, and D under Series III are greater than the means for Series II. The mean of Group A decreased slightly but not significantly so. The reliabilities of the differences for the other three groups are statistically significant in two cases and in the third approaches significance. It seems highly probable that failure-repetition, failure-reproof, and failure-punishment are more effective for performance than are success-repetition, success-praise, and success-reward. However, these results on Series III may have been influenced by at least two factors: (1) greater interest in the train system than in the water system, or (2) having had a previous success with the same experimenter. Until these two factors have been investigated it is only possible to conclude that if a child has had a previous success with an experimenter and returned back one week later to perform a different task in which he might be interested to a greater extent than in the previous one, he will tend to expend a greater amount of energy under failure than he did under success.

7. Motivation is not operating to inhibit variability since there were no significant differences between the coefficients of variation except where the differences might be covered up because of correlation between the means and the standard deviations. In the latter cases where both increased significantly, they also increased proportionately.

8. There was a greater percentage of subjects who in any seven trials given on one day decreased their scores in the seventh trial over the second trial than the percentage of subjects who increased or remained the same as in first trial. The percentage of subjects who decreased 10 per cent or more, 15 per cent or more, 25 per cent or more was greater under control-motivation than under success-repetition, success-praise, success-reward, failure-repetition, failure-reproof, or failure-punishment. It may be concluded that the types of motivation given not only increase the output of energy but also tend to maintain it to a greater extent than does control-motivation.

9. The qualitative aspects of the influence of motivation may be noted as:

- (a) The presence or absence of seventeen items under expressive behavior was checked by a recorder for each child during each of the seven trials given on any one day. The maximum score for all items except 1 and 7 was seven for any one test series. The relative frequencies of occurrence of these were computed together with their probable errors and the reliabilities of the differences. Items 1 and 7 were rated on a basis of three points and the maximum score for these items was, therefore, twenty-one. The mean probable errors of the means and the standard deviations of these items were computed.
- (b) Results show that one influence of motivation is a tendency to increase the amount of energy output that is directed through other channels (neck, face, etc.) than through the main channel (the hand and arm).
- (c) The percent of subjects liking Game I or II varies according to groups and series. In general, it may be stated that in this study a significantly larger percentage of subjects preferred the game with some motivation to one that was given under control-motivation, even though they were not successful in attaining the goal set.
- (d) In general, there is a tendency for the child to be rated as being elated under success, while under failure there is a tendency for subject to be rated as "depressed" on Mood 2. There were no significant differences for rating given in Mood 1. Under Mood 3 there was some slight indication that a child tended to be rated as more depressed after failure. This tendency approaches significance for two groups and is significant in the other group.

10. There was a tendency for a greater percentage of younger children to increase their scores under Series II and III than older ones. The reason for the lower initial scores for the younger subjects is not clear. There were no consistent sex differences. These data are inadequate, however, for drawing any conclusions concerning the relationship of either age or sex to motivation.

11. It may be concluded that although there were significant differences between test series (in the practice effects) on the Perforation Test there seemed to be very little indication that there

were carry-over effects due to the different types of incentives used in a previous task.

The reliability of the Perforation Test for the fifty subjects given three applications of the test each one week apart varied from $.887 \pm .02$ to $.90 \pm .017$. The coefficient obtained from a single application of the test for 211 children ranging in age from twenty-seven to ninety-six months varied according to the trial from $.92 \pm .007$ to $.94 \pm .006$.

REFERENCES

1. Alpert, Augusta: *The solving of problem-situations by preschool children: An analysis.* Teach. Coll., Columbia Univ. Cont. to Educ., 1928, No. 323, Pp. 69.
2. Arps, George F.: Work with knowledge of results versus work without knowledge of results: Awareness and partial awareness as factors conditioning efficiency. *Psychol. Rev., Monog. Suppl.*, 1920, 28, No. 3, Pp. 41.
3. Baldwin, Bird T., and Stecher, Lorle I.: *The psychology of the pre-school child.* New York: Appleton, 1924. Pp. 305.
4. Bird, Grace E.: *Personality factors in learning.* Person. J., 1927-1928, 6, 56-59.
5. Blanchard, Phyllis M.: *The child and society: An introduction to the social psychology of the child.* New York: Longmans, Green, 1928. Pp. xi, 569.
6. Blanton, Smiley, and Blanton, Margaret Gray: *Child guidance.* New York: Century, 1927. Pp. xviii, 301.
7. Blatz, William E.: Discipline versus corporal punishment. *Childhood Educ.*, 1928-1929, 5, 144-149.
8. Book, William E., and Norvell, Lee: The will to learn: An experimental study of incentives in learning. *Ped. Sem.*, 1922, 29, 305-362.
9. Briggs, Thomas H.: Praise and censure as incentives. *School & Soc.*, 1927, 26, 596-598.
10. Bronner, Augusta F.: Attitude as it affects performance of tests. *Psychol. Rev.*, 1916, 23, 303-331.
11. Chapman, J. Crosby, and Feder, R. B.: The effect of external incentives on improvement. *J. Educ. Psychol.*, 1917, 8, 469-474.
12. Cole, Lawrence W.: The relation of strength of stimulus to rate of learning in the chick. *J. Animal Behavior*, 1911, 1, 111-124.
13. Dodson, J. D.: Relative values of reward and punishment in habit formation. *Psychobiol.*, 1917, 1, 231-276.
14. Dunlap, Knight, and Snyder, Agnes: Practice effects in intelligence tests. *J. Exper. Psychol.*, 1920, 3, 396-403.
15. English, Horace B.: *A student's dictionary of psychological terms.* Yellow Springs, Ohio: Antioch Press, 1928. [Pp. 79].
16. Freeman, Frank N.: How children learn. *Child Study*, 1929-1930, 7, 33-35.
17. Garrett, Henry E.: *Statistics in psychology and education.* New York: Longmans, Green, 1926. Pp. xiii, 317.
18. Gates, Georgina Stickland, and Rissland, Louise Q.: The effect of encouragement and of discouragement upon performance. *J. Educ. Psychol.*, 1923, 14, 21-26.
19. Gilchrist, Edward P.: The extent to which praise and reproof affect a pupil's work. *School & Soc.*, 1916, 4, 872-874.

20. Goodenough, Florence L., and Brian, Clara R.: Certain factors underlying the acquisition of motor skill by pre-school children. *J. Exper. Psychol.*, 1929, 12, 127-155.
21. Haggerty, M. E.: The incidence of undesirable behavior in public-school children. *J. Educ. Res.*, 1925, 12, 102-122.
22. Hamilton, Hughbert C.: The effect of incentives on accuracy of discrimination measured on the Galton bar. *Arch. Psychol.*, 1928-1929, 16, No. 103, Pp. 73.
23. Hicks, James Allan: *The acquisition of motor skill in young children: An experimental study of the effects of practice in throwing at a moving target.* Univ. Iowa Stud., Stud. in Child Welfare, 1931, 4, No. 6, Pp. 80.
24. Hoge, Mildred A., and Stocking, Ruth J.: A note on the relative value of punishment and reward as motives. *J. Animal Behavior*, 1912, 2, 43-50.
25. Holzinger, Karl J.: *Statistical methods for students in education.* New York: Ginn, 1928. Pp. viii, 372.
26. Hurlock, Elizabeth B.: The value of praise and reproof as incentives for children. *Arch. Psychol.*, 1924-1925, 11, No. 71, 5-78.
27. Hurlock, Elizabeth B.: The effect of incentives upon the constancy of the I. Q. *Ped. Sem.*, 1925, 32, 422-434.
28. Johanson, Albert M.: The influence of incentive and punishment upon reaction time. *Arch. Psychol.*, 1922, 8, No. 54, Pp. 52 (p. 5-52)
29. Johnson, Buford J.: *Mental growth of children in relation to rate of growth in bodily development.* New York: E. P. Dutton, 1925. Pp. xix, 160.
30. Judd, Charles H.: Practice without knowledge of results. [In] Yale Psychological Studies, New Series, Vol. I, Pp. vii, 226. *Psychol. Rev., Monog. Suppl.*, 1905, 7, No. 1, 185-194.
31. Kilpatrick, William H.: Behavior problems. *Childhood Educ.*, 1928, 5, 119-126.
32. Kirby, Thomas Joseph: *Practice in the case of school children.* Teach. Coll., Columbia Univ. Cont. to Educ., 1913, No. 58, Pp. 98.
33. Knight, Frederic B., and Remmers, H. H.: Fluctuations in mental production when motivation is the main variable. *J. Appl. Psychol.*, 1923, 7, 209-223.
34. Lowrey, Lawson G.: Parental conflicts and their effects on the personality of children. *Child Study*, 1929-1930, 7, 1-3.
35. Merchant, F. W., and Chant, C. A.: *Elements of physics.* New York: Henry Holt, 1924. Pp. 546.
36. Metfessel, Milton: The strobophotograph: A device for measuring pitch. *J. Gen. Psychol.*, 1929, 2, 135-139.
37. Moore, Elizabeth S.: *The development of mental hygiene in young children through purposeful activity.* Univ. Iowa Stud., Stud. in Child Welfare, 1931, 4, No. 2, Pp. 128.
38. Murray, Sir James [Editor]: *New English dictionary on historical principles: Founded mainly on the materials collected by the Philological Society.* 10 Vols., Oxford: Clarendon Press, 1908. [Pp. 18, 142]

39. Nelson, Janet Fowler: *A preliminary report on some uses of the psychological test situation for studying personality difference*. [In] Thomas, Dorothy S. and associates: *Some New Techniques for Studying Social Behavior*. Teach. Coll., Columbia Univ., *Child Development Monog.*, 1929, No. 1, Pp. x, 203.
40. Rexroad, Carl N.: Administering electric shock for inaccuracy in multiple choice reactions. *J. Exper. Psychol.*, 1926, 9, 1-18.
41. Ross, Clay Campbell: An experiment in motivation. *J. Educ. Psychol.*, 1927, 18, 337-346.
42. Ruger, Henry Alford: The psychology of efficiency: An experimental study of the processes involved in the solution of mechanical puzzles and in the acquisition of skill in their manipulation. *Arch. Psychol.*, 1910, 19, No. 15, Pp. 88.
43. Spencer, Llewellyn T.: The effect of practice without knowledge of results. *Amer. J. Psychol.*, 1923, 34, 107-111.
44. Sullivan, Ellen B.: Attitude in relation to learning. *Psychol. Rev. Monog.*, 1927, 36, No. 3, Pp. 149.
45. Terman, Lewis M.: *The measurement of intelligence: An explanation of and a complete guide for the use of the Stanford Revision and extension of the Binet-Simon Intelligence Scale*. New York: Houghton Mifflin, 1916. Pp. xviii, 362.
46. Thom, Douglas A.: *Every-day problems of the every-day child*. New York: Appleton, 1929. Pp. xiii, 349.
47. Thomson, Mehran K.: *The springs of human actions: A psychological study of the sources, mechanism, and principles of motivation in human behavior*. New York: Appleton, 1927. Pp. xvi, 501.
48. Thorndike, Edward L.: *Educational Psychology. Vol. II. The Psychology of Learning*. New York: Teach. Coll., Columbia Univ., 1913. Pp. xi, 452.
49. Thorndike, Edward L.: The law of effect. *Amer. J. Psychol.*, 1927, 39, 212-222.
50. Thorndike, Edward L.: *Fundamental factors in learning*. [In] *Proceedings and Papers of the Ninth International Congress of Psychology*. Held at Yale University, New Haven, Conn., September 1 to 7, 1929. Princeton: Psychological Review Co., [c. 1930] Pp. 534 (p. 429)
51. Thorndike, Edward L.: *The strengthening and weakening of mental connections by their after-effects*. [In] *Proceedings and Papers of the Ninth International Congress of Psychology*. Held at Yale University, New Haven, Conn., September 1 to 7, 1929. Princeton: Psychological Review Co., [c. 1930] Pp. 534 (p. 430)
52. Warden, Carl J., and Aylesworth, Mercy: The relative value of reward and punishment in the formation of a visual discrimination habit in the white rat. *J. Comp. Psychol.*, 1927, 7, 117-127.
53. Wickman, E. K.: *Children's behavior and teachers' attitudes*. New York: Commonwealth Fund Division of Publications, 1928. Pp. 247.
54. Woodworth, Robert S.: *Psychology: A study of mental life*. New York: Henry Holt, 1921. Pp. x, 580.

55. Wright, William R.: Some effects of incentives on work and fatigue.
Psychol. Rev., 1906, 13, 23-34.
56. Yerkes, Robert M.: *The dancing mouse: A study in animal behavior*.
New York: Macmillan, 1907. Pp. xxi, 290.
57. Yerkes, Robert M., and Dodson, John D.: The relation of strength of
stimulus to rapidity of habit-formation. *J. Comp. Neurol. & Psychol.*,
1908, 18, 459-482.

LB1105 .I64 v.5:3
Motivation of young children; an

Princeton Theological Seminary-Speer Library



1 1012 00141 8195